1. Preface Optics
2. The Nature of Light
    1. Introduction
    2. The Propagation of Light
    3. The Law of Reflection
    4. Refraction
    5. Total Internal Reflection
    6. Dispersion
    7. Huygens’s Principle
    8. Polarization
3. Geometric Optics and Image Formation
    1. Introduction
    2. Images Formed by Plane Mirrors
    3. Spherical Mirrors
    4. Images Formed by Refraction
    5. Thin Lenses
    6. The Eye
    7. The Camera
    8. The Simple Magnifier
    9. Microscopes and Telescopes
4. Waves
    1. Traveling Waves
    2. Mathematics of Waves
    3. Energy and Power of a Wave
    4. Interference of Waves
    5. Standing Waves and Resonance
5. Interference
    1. Introduction
    2. Young's Double-Slit Interference
    3. Mathematics of Interference
    4. Multiple-Slit Interference

    5. Interference in Thin Films
    6. The Michelson Interferometer
6. Diffraction
    1. Introduction
    2. Single-Slit Diffraction
    3. Intensity in Single-Slit Diffraction
    4. Double-Slit Diffraction
    5. Diffraction Gratings
    6. Circular Apertures and Resolution
    7. X-Ray Diffraction
    8. Holography
7. Phasors
    1. Phasors: Phasor Representation of Signals
    2. Phasors: Introduction
    3. Phasors: Beating between Tones
    4. Phasors: Multiphase Power
    5. Phasors: Light Scattering by a Slit
8. Fundamental Constants
9. Mathematical Formulas

Preface Optics

This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

About OpenStax

OpenStax is a nonprofit based at Rice University, and it’s our mission to improve student access to education. Our first openly licensed college textbook was published in 2012 and our library has since scaled to over 20 books used by hundreds of thousands of students across the globe. Our adaptive learning technology is designed to improve learning outcomes through personalized educational paths. The OpenStax mission is made possible through the generous support of philanthropic foundations. Through these partnerships and with the help of additional low-cost resources from our OpenStax partners, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

About *Optics*

Coverage and Scope

Chapter 1: The Nature of Light
Chapter 2: Geometric Optics and Image Formation
Chapter 3: Interference
Chapter 4: Diffraction

Pedagogical Foundation

Throughout this textbook, you will find derivations of concepts that present classical ideas and techniques, as well as modern applications and methods. Most chapters start with observations or experiments that place the material in a context of physical experience. Presentations and explanations rely on years of classroom experience on the part of long-time physics professors, striving for a balance of clarity and rigor that has proven successful with their students. Throughout the text, links enable students to review earlier material and then return to the present discussion, reinforcing connections between topics. Key historical figures and experiments are discussed in the main text (rather than in boxes or sidebars), maintaining a focus on the development of physical intuition. Key ideas, definitions, and equations are highlighted in the text and listed in summary form at the end of each chapter. Examples and chapter-opening images often include contemporary applications from daily life or modern science and engineering.

Assessments That Reinforce Key Concepts

In-chapter **Examples** generally follow a three-part format of Strategy, Solution, and Significance to emphasize how to approach a problem, how to work with the equations, and how to check and generalize the result. Examples are often followed by **Check Your Understanding** questions and answers to help reinforce for students the important ideas of the examples. **Problem-Solving Strategies** in each chapter break down methods of approaching various types of problems into steps students can follow for guidance. The book also includes exercises at the end of each chapter so students can practice what they’ve learned.

> **Conceptual questions** do not require calculation but test student learning of the key concepts.
> **Problems** categorized by section test student problem-solving skills and the ability to apply ideas to practical situations.
> **Additional Problems** apply knowledge across the chapter, forcing students to identify what concepts and equations are appropriate for solving given problems. Randomly located throughout the problems are **Unreasonable Results** exercises that ask students to evaluate the answer to a problem and explain why it is not reasonable and what assumptions made might not be correct.
> **Challenge Problems** extend text ideas to interesting but difficult situations.

Answers for selected exercises are available in an **Answer Key** at the end of the book.

About the Authors

Senior Contributing Authors

Samuel J. Ling, Truman State University

Jeff Sanny, Loyola Marymount University

Bill Moebs, PhD

Introduction class="introduction"

Due to total internal reflection, an underwater swimmer’s image is reflected back into the water where the camera is located. The circular ripple in the image center is actually on the water surface. Due to the viewing angle, total internal reflection is not occurring at the top edge of this image, and we can see a view of activities on the pool deck. (credit: modification of work by “jayhem”/Flickr)

Our investigation of light revolves around two questions of fundamental importance: (1) What is the nature of light, and (2) how does light behave under various circumstances? Answers to these questions can be found in Maxwell’s equations (in Electromagnetic Waves), which predict the existence of electromagnetic waves and their behavior. Examples of light include radio and infrared waves, visible light, ultraviolet radiation, and X-rays. Interestingly, not all light phenomena can be explained by Maxwell’s theory. Experiments performed early in the twentieth century showed that light has corpuscular, or particle-like, properties. The idea that light can display both wave and particle characteristics is called *wave-particle duality*, which is examined in Photons and Matter Waves.

In this chapter, we study the basic properties of light. In the next few chapters, we investigate the behavior of light when it interacts with optical devices such as mirrors, lenses, and apertures.

The Propagation of Light

By the end of this section, you will be able to:

- Determine the index of refraction, given the speed of light in a medium
- List the ways in which light travels from a source to another location

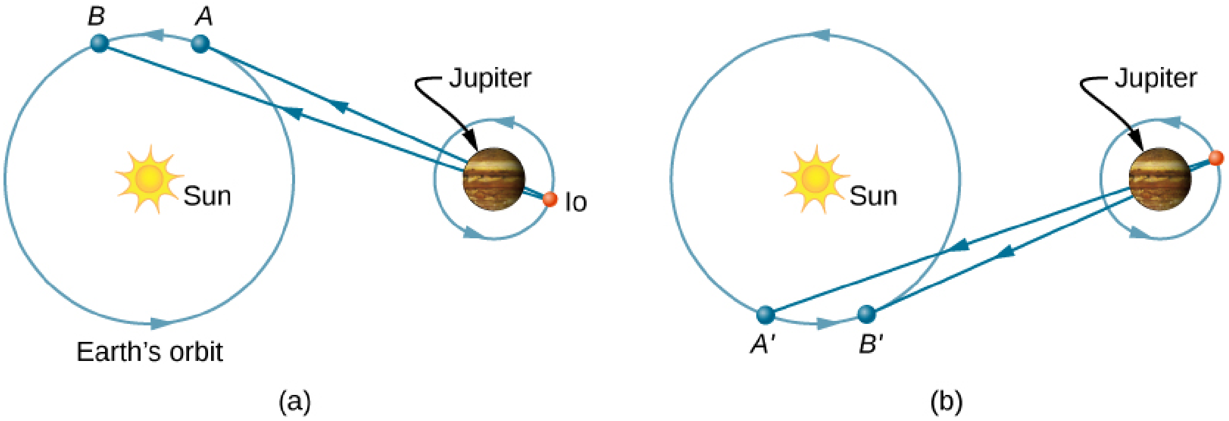
The speed of light in a vacuum c is one of the fundamental constants of physics. As you will see when you reach Relativity, it is a central concept in Einstein’s theory of relativity. As the accuracy of the measurements of the speed of light improved, it was found that different observers, even those moving at large velocities with respect to each other, measure the same value for the speed of light. However, the speed of light does vary in a precise manner with the material it traverses. These facts have far-reaching implications, as we will see in later chapters.

The Speed of Light: Early Measurements

The first measurement of the speed of light was made by the Danish astronomer Ole Roemer (1644–1710) in 1675. He studied the orbit of Io, one of the four large moons of Jupiter, and found that it had a period of revolution of 42.5 h around Jupiter. He also discovered that this value fluctuated by a few seconds, depending on the position of Earth in its orbit around the Sun. Roemer realized that this fluctuation was due to the finite speed of light and could be used to determine c.

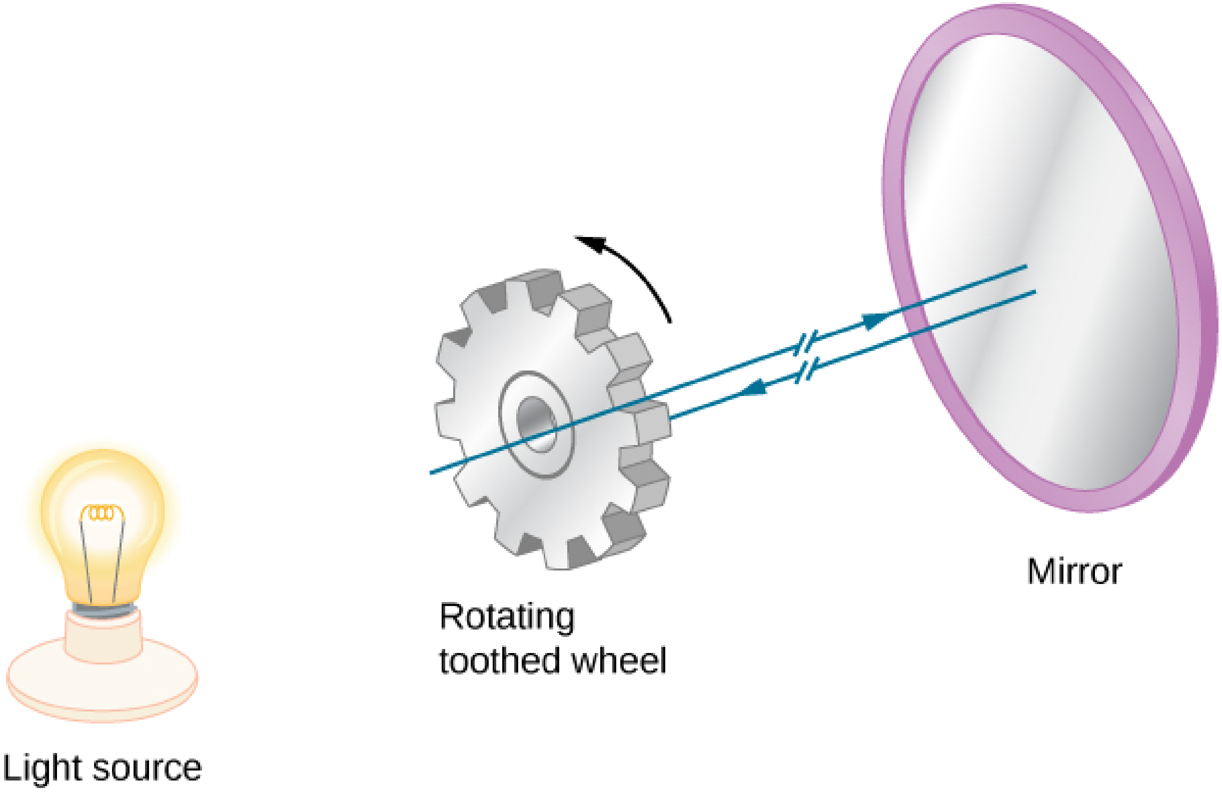
Roemer found the period of revolution of Io by measuring the time interval between successive eclipses by Jupiter. [link](a) shows the planetary configurations when such a measurement is made from Earth in the part of its orbit where it is receding from Jupiter. When Earth is at point A, Earth, Jupiter, and Io are aligned. The next time this alignment occurs, Earth is at point B, and the light carrying that information to Earth must travel to that point. Since B is farther from Jupiter than A, light takes more time to reach Earth when Earth is at B. Now imagine it is about 6 months later, and the planets are arranged as in part (b) of the figure. The measurement of Io’s period begins with Earth at point A' and Io eclipsed by Jupiter. The next eclipse then occurs when Earth is at point B', to which the light carrying the information of this eclipse must travel. Since B' is closer to Jupiter than

A', light takes less time to reach Earth when it is at B'. This time interval between the successive eclipses of Io seen at A' and B' is therefore less than the time interval between the eclipses seen at A and B. By measuring the difference in these time intervals and with appropriate knowledge of the distance between Jupiter and Earth, Roemer calculated that the speed of light was $2.0 \times 10^8\,\text{m/s}$, which is 33% below the value accepted today.



Roemer's astronomical method for determining the speed of light. Measurements of Io's period done with the configurations of parts (a) and (b) differ, because the light path length and associated travel time increase from A to B (a) but decrease from A' to B' (b).

The first successful terrestrial measurement of the speed of light was made by Armand Fizeau (1819–1896) in 1849. He placed a toothed wheel that could be rotated very rapidly on one hilltop and a mirror on a second hilltop 8 km away ([link]). An intense light source was placed behind the wheel, so that when the wheel rotated, it chopped the light beam into a succession of pulses. The speed of the wheel was then adjusted until no light returned to the observer located behind the wheel. This could only happen if the wheel rotated through an angle corresponding to a displacement of $(n + ½)$ teeth, while the pulses traveled down to the mirror and back. Knowing the rotational speed of the wheel, the number of teeth on the wheel, and the distance to the mirror, Fizeau determined the speed of light to be $3.15 \times 10^8\,\mathrm{m/s}$, which is only 5% too high.



Fizeau's method for measuring the speed of light. The teeth of the wheel block the reflected light upon return when the wheel is rotated at a rate that matches the light travel time to and from the mirror.

The French physicist Jean Bernard Léon Foucault (1819–1868) modified Fizeau's apparatus by replacing the toothed wheel with a rotating mirror. In 1862, he measured the speed of light to be $2.98 \times 10^8\,\mathrm{m/s}$, which is within 0.6% of the presently accepted value. Albert Michelson (1852–1931) also used Foucault's method on several occasions to measure the speed of light. His first experiments were performed in 1878; by 1926, he had refined the technique so well that he found *c* to be $(2.99796 \pm 4) \times 10^8\,\mathrm{m/s}$.

Today, the speed of light is known to great precision. In fact, the speed of light in a vacuum *c* is so important that it is accepted as one of the basic physical quantities and has the value

Note:
Equation:

$$c = 2.99792458 \times 10^{8}\,\mathrm{m/s} \approx 3.00 \times 10^{8}\,\mathrm{m/s}$$

where the approximate value of $3.00 \times 10^{8}\,\mathrm{m/s}$ is used whenever three-digit accuracy is sufficient.

Speed of Light in Matter

The speed of light through matter is less than it is in a vacuum, because light interacts with atoms in a material. The speed of light depends strongly on the type of material, since its interaction varies with different atoms, crystal lattices, and other substructures. We can define a constant of a material that describes the speed of light in it, called the **index of refraction** *n*:

Note:
Equation:

$$n = \frac{c}{v}$$

where *v* is the observed speed of light in the material.

Since the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction is always greater than or equal to one; that is, $n \geq 1$. [link] gives the indices of refraction for some representative substances. The values are listed for a particular wavelength of light, because they vary slightly with wavelength. (This can have important effects, such as colors separated by a prism, as we will see in Dispersion.) Note that for gases, n is close to 1.0. This seems reasonable, since atoms in gases are widely separated, and light travels at c in the vacuum between atoms. It is common to take $n = 1$ for gases unless great precision is needed. Although the speed of light v in a medium varies considerably from its value c in a vacuum, it is still a large speed.

Medium	***n***
Gases at 0°C, 1 atm	
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
Liquids at 20°C	
Benzene	1.501
Carbon disulfide	1.628
Carbon tetrachloride	1.461

Medium	***n***
Ethanol	1.361
Glycerine	1.473
Water, fresh	1.333
Solids at 20°C	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice (at 0°C)	1.309
Polystyrene	1.49
Plexiglas	1.51
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544
Zircon	1.923

Index of Refraction in Various MediaFor light with a wavelength of 589 nm in a vacuum

Example:

Speed of Light in Jewelry

Calculate the speed of light in zircon, a material used in jewelry to imitate diamond.

Strategy

We can calculate the speed of light in a material *v* from the index of refraction *n* of the material, using the equation $n = c/v$.

Solution

Rearranging the equation $n = c/v$ for *v* gives us

Equation:

$$v = \frac{c}{n}.$$

The index of refraction for zircon is given as 1.923 in [link], and *c* is given in [link]. Entering these values in the equation gives

Equation:

$$v = \frac{3.00 \times 10^8 \, \text{m/s}}{1.923} = 1.56 \times 10^8 \, \text{m/s}.$$

Significance

This speed is slightly larger than half the speed of light in a vacuum and is still high compared with speeds we normally experience. The only substance listed in [link] that has a greater index of refraction than zircon is diamond. We shall see later that the large index of refraction for zircon makes it sparkle more than glass, but less than diamond.

Note:

Exercise:

Problem:

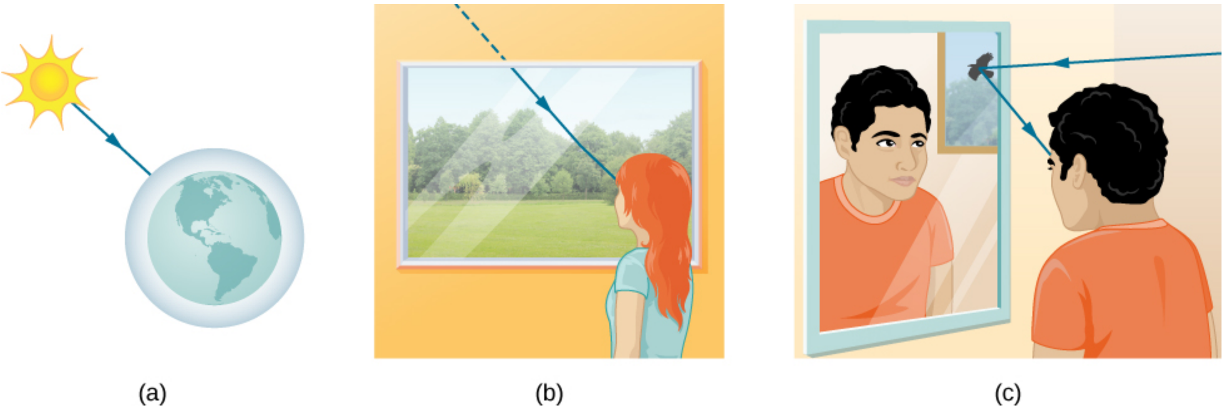
Check Your Understanding [link] shows that ethanol and fresh water have very similar indices of refraction. By what percentage do the speeds of light in these liquids differ?

Solution:

2.1% (to two significant figures)

The Ray Model of Light

You have already studied some of the wave characteristics of light in the previous chapter on Electromagnetic Waves. In this chapter, we start mainly with the ray characteristics. There are three ways in which light can travel from a source to another location ([link]). It can come directly from the source through empty space, such as from the Sun to Earth. Or light can travel through various media, such as air and glass, to the observer. Light can also arrive after being reflected, such as by a mirror. In all of these cases, we can model the path of light as a straight line called a **ray**.



Three methods for light to travel from a source to another location. (a) Light reaches the upper atmosphere of Earth, traveling through empty space directly from the source. (b) Light can reach a person by traveling through media like air and glass. (c) Light can also reflect from an object like a mirror. In the situations shown here, light interacts with objects large enough that it travels in straight lines, like a ray.

Experiments show that when light interacts with an object several times larger than its wavelength, it travels in straight lines and acts like a ray. Its wave characteristics are not pronounced in such situations. Since the wavelength of visible light is less than a micron (a thousandth of a millimeter), it acts like a ray in the many common situations in which it encounters objects larger than a micron. For example, when visible light encounters anything large enough that we can observe it with unaided eyes, such as a coin, it acts like a ray, with generally negligible wave characteristics.

In all of these cases, we can model the path of light as straight lines. Light may change direction when it encounters objects (such as a mirror) or in passing from one material to another (such as in passing from air to glass), but it then continues in a straight line or as a ray. The word “ray” comes from mathematics and here means a straight line that originates at some point. It is acceptable to visualize light rays as laser rays. The *ray model of light* describes the path of light as straight lines.

Since light moves in straight lines, changing directions when it interacts with materials, its path is described by geometry and simple trigonometry. This part of optics, where the ray aspect of light dominates, is therefore called **geometric optics**. Two laws govern how light changes direction when it interacts with matter. These are the *law of reflection*, for situations in which light bounces off matter, and the *law of refraction*, for situations in which light passes through matter. We will examine more about each of these laws in upcoming sections of this chapter.

Summary

- The speed of light in a vacuum is $c = 2.99792458 \times 10^8 \, \text{m/s} \approx 3.00 \times 10^8 \, \text{m/s}.$
- The index of refraction of a material is $n = c/v$, where v is the speed of light in a material and c is the speed of light in a vacuum.
- The ray model of light describes the path of light as straight lines. The part of optics dealing with the ray aspect of light is called geometric optics.

- Light can travel in three ways from a source to another location: (1) directly from the source through empty space; (2) through various media; and (3) after being reflected from a mirror.

Conceptual Questions

Exercise:

Problem:

Under what conditions can light be modeled like a ray? Like a wave?

Solution:

Light can be modeled as a ray when devices are large compared to wavelength, and as a wave when devices are comparable or small compared to wavelength.

Exercise:

Problem:

Why is the index of refraction always greater than or equal to 1?

Exercise:

Problem:

Does the fact that the light flash from lightning reaches you before its sound prove that the speed of light is extremely large or simply that it is greater than the speed of sound? Discuss how you could use this effect to get an estimate of the speed of light.

Solution:

This fact simply proves that the speed of light is greater than that of sound. If one knows the distance to the location of the lightning and the speed of sound, one could, in principle, determine the speed of light from the data. In practice, because the speed of light is so great, the data would have to be known to impractically high precision.

Exercise:

Problem:

Speculate as to what physical process might be responsible for light traveling more slowly in a medium than in a vacuum.

Problems

Exercise:

Problem: What is the speed of light in water? In glycerine?

Exercise:

Problem: What is the speed of light in air? In crown glass?

Solution:

$2.99705 \times 10^8 \text{ m/s}$; $1.97 \times 10^8 \text{ m/s}$

Exercise:

Problem:

Calculate the index of refraction for a medium in which the speed of light is $2.012 \times 10^8 \text{ m/s}$, and identify the most likely substance based on [link].

Exercise:

Problem:

In what substance in [link] is the speed of light $2.290 \times 10^8 \text{ m/s}$?

Solution:

ice at 0°C

Exercise:

Problem:

There was a major collision of an asteroid with the Moon in medieval times. It was described by monks at Canterbury Cathedral in England as a red glow on and around the Moon. How long after the asteroid hit the Moon, which is $3.84 \times 10^5 \text{ km}$ away, would the light first arrive on Earth?

Exercise:

Problem:

Components of some computers communicate with each other through optical fibers having an index of refraction $n = 1.55$. What time in nanoseconds is required for a signal to travel 0.200 m through such a fiber?

Solution:

1.03 ns

Exercise:

Problem:

Compare the time it takes for light to travel 1000 m on the surface of Earth and in outer space.

Exercise:

Problem:

How far does light travel underwater during a time interval of 1.50×10^{-6} s?

Solution:

337 m

Glossary

geometric optics
part of optics dealing with the ray aspect of light

index of refraction
for a material, the ratio of the speed of light in a vacuum to that in a material

ray
straight line that originates at some point

The Law of Reflection

By the end of this section, you will be able to:

- Explain the reflection of light from polished and rough surfaces
- Describe the principle and applications of corner reflectors

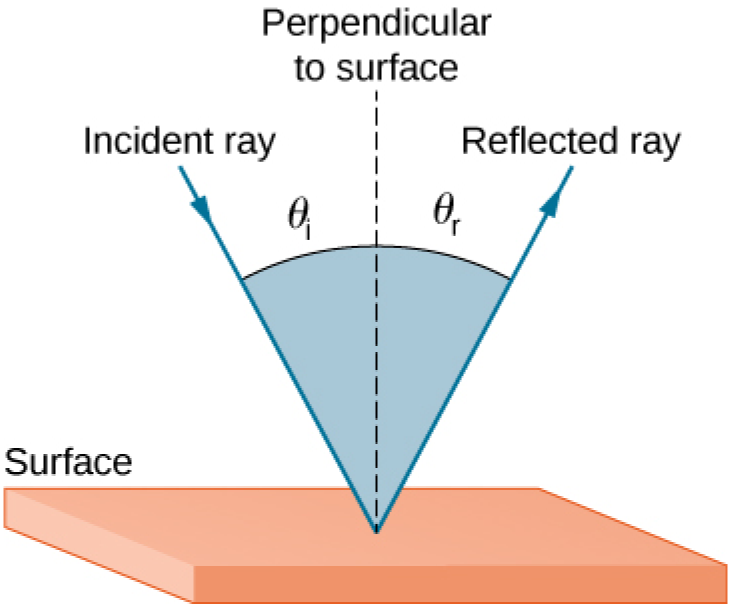
Whenever we look into a mirror, or squint at sunlight glinting from a lake, we are seeing a reflection. When you look at a piece of white paper, you are seeing light scattered from it. Large telescopes use reflection to form an image of stars and other astronomical objects.

The **law of reflection** states that the angle of reflection equals the angle of incidence, or

Note:
Equation:

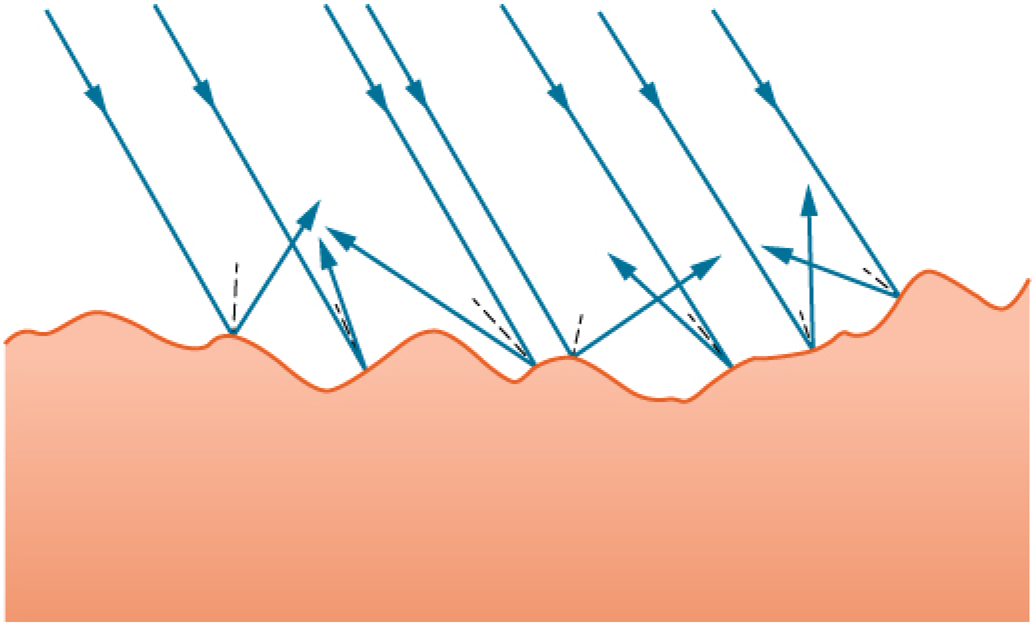
$$\theta_{\text{r}} = \theta_{\text{i}}$$

The law of reflection is illustrated in [link], which also shows how the angle of incidence and angle of reflection are measured relative to the perpendicular to the surface at the point where the light ray strikes.

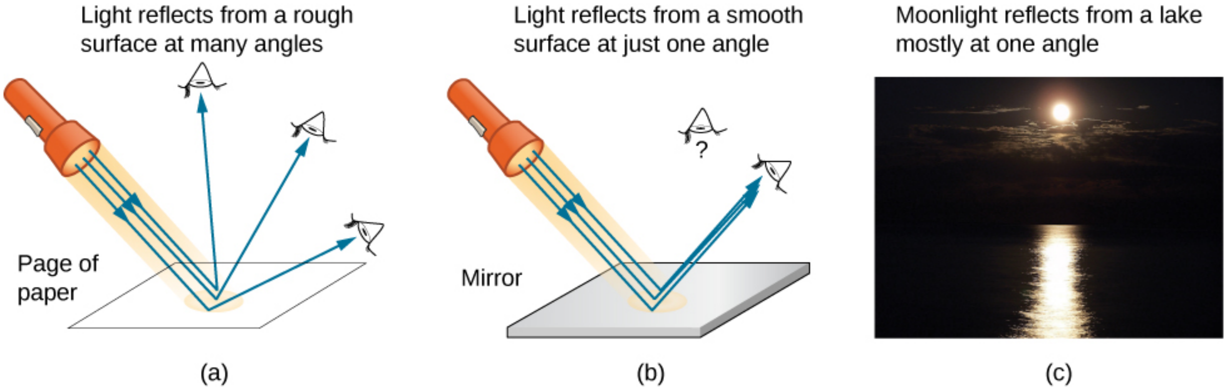


The law of reflection states that the angle of reflection equals the angle of incidence—$\theta_r = \theta_i$. The angles are measured relative to the perpendicular to the surface at the point where the ray strikes the surface.

We expect to see reflections from smooth surfaces, but [link] illustrates how a rough surface reflects light. Since the light strikes different parts of the surface at different angles, it is reflected in many different directions, or diffused. Diffused light is what allows us to see a sheet of paper from any angle, as shown in [link](a). People, clothing, leaves, and walls all have rough surfaces and can be seen from all sides. A mirror, on the other hand, has a smooth surface (compared with the wavelength of light) and reflects light at specific angles, as illustrated in [link](b). When the Moon reflects from a lake, as shown in [link](c), a combination of these effects takes place.

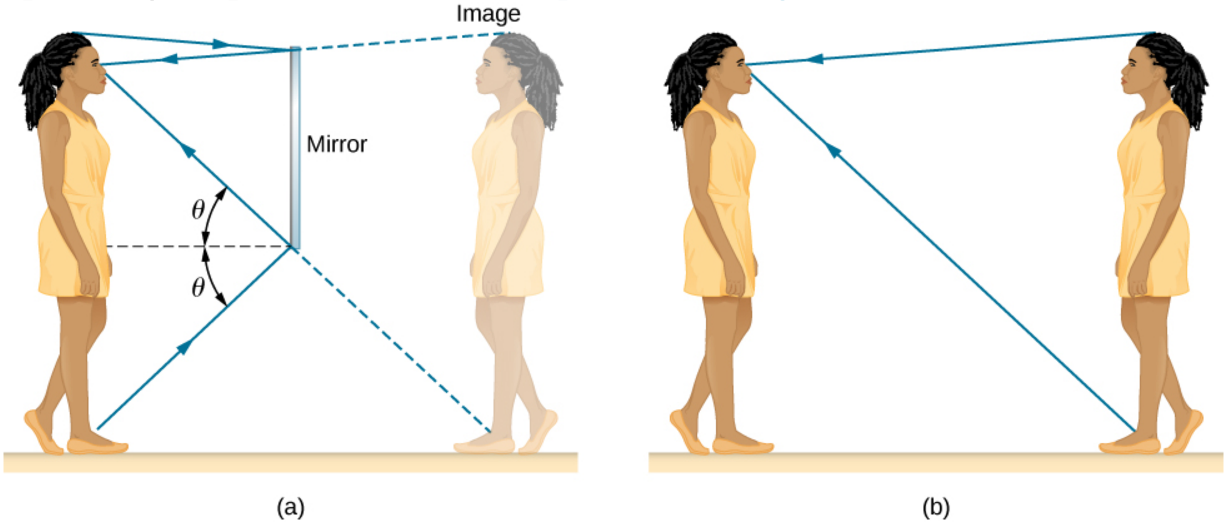


Light is diffused when it reflects from a rough surface. Here, many parallel rays are incident, but they are reflected at many different angles, because the surface is rough.


Light reflects from a rough surface at many angles
Page of paper
(a)
Light reflects from a smooth surface at just one angle
?
Mirror
(b)
Moonlight reflects from a lake mostly at one angle
(c)

(a) When a sheet of paper is illuminated with many parallel incident rays, it can be seen at many different angles, because its surface is rough and diffuses the light. (b) A mirror illuminated by many parallel rays reflects them in only one direction, because its surface is very smooth. Only the observer at a particular angle sees the reflected light. (c) Moonlight is spread out when it is reflected by the lake, because the surface is shiny but uneven. (credit c: modification of work by Diego Torres Silvestre)

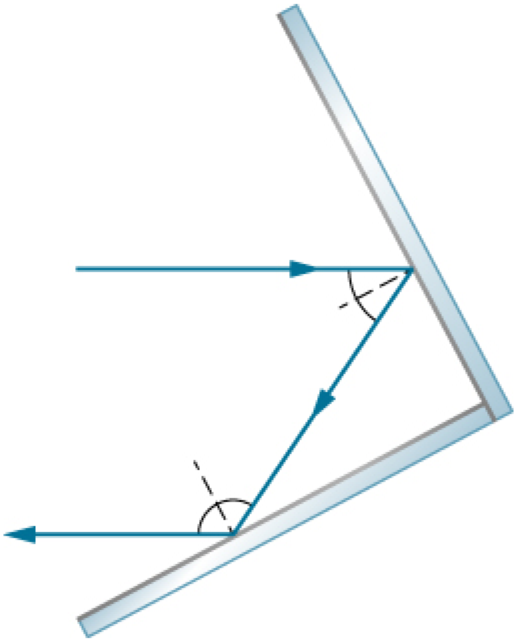
When you see yourself in a mirror, it appears that the image is actually behind the mirror ([link]). We see the light coming from a direction determined by the law of reflection. The angles are such that the image is exactly the same distance behind the mirror as you stand in front of the mirror. If the mirror is on the wall of a room, the images in it are all behind the mirror, which can make the room seem bigger. Although these mirror images make objects appear to be where they cannot be (like behind a solid wall), the images are not figments of your imagination. Mirror images can be photographed and videotaped by instruments and look just as they do with our eyes (which are optical instruments themselves). The precise manner in which images are formed by mirrors and lenses is discussed in an upcoming chapter on Geometric Optics and Image Formation.



(a) Your image in a mirror is behind the mirror. The two rays shown are those that strike the mirror at just the correct angles to be reflected into the eyes of the person. The image appears to be behind the mirror at the same distance away as (b) if you were looking at your twin directly, with no mirror.

Corner Reflectors (Retroreflectors)

A light ray that strikes an object consisting of two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came ([link]). This is true whenever the reflecting surfaces are perpendicular, and it is independent of the angle of incidence. (For proof, see [link] at the end of this section.) Such an object is called a **corner reflector**, since the light bounces from its inside corner. Corner reflectors are a subclass of retroreflectors, which all reflect rays back in the directions from which they came. Although the geometry of the proof is much more complex, corner reflectors can also be built with three mutually perpendicular reflecting surfaces and are useful in three-dimensional applications.



A light ray that strikes two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came.

Many inexpensive reflector buttons on bicycles, cars, and warning signs have corner reflectors designed to return light in the direction from which it originated. Rather than simply reflecting light over a wide angle, retroreflection ensures high visibility if the observer and the light source are located together, such as a car’s driver and headlights. The Apollo astronauts placed a true corner reflector on the Moon ([link]). Laser signals from Earth can be bounced from that corner reflector to measure the gradually increasing distance to the Moon of a few centimeters per year.

(a) Astronauts placed a corner reflector on the Moon to measure its gradually increasing orbital distance. (b) The bright spots on these bicycle safety reflectors are reflections of the flash of the camera that took this picture on a dark night. (credit a: modification of work by NASA; credit b: modification of work by “Julo”/Wikimedia Commons)

Working on the same principle as these optical reflectors, corner reflectors are routinely used as radar reflectors ([link]) for radio-frequency applications. Under most circumstances, small boats made of fiberglass or wood do not strongly reflect radio waves emitted by radar systems. To make these boats visible to radar (to avoid collisions, for example), radar reflectors are attached to boats, usually in high places.

A radar reflector hoisted on a sailboat is a type of corner reflector. (credit: Tim Sheerman-Chase)

As a counterexample, if you are interested in building a stealth airplane, radar reflections should be minimized to evade detection. One of the design considerations would then be to avoid building 90° corners into the airframe.

Summary

- When a light ray strikes a smooth surface, the angle of reflection equals the angle of incidence.
- A mirror has a smooth surface and reflects light at specific angles.
- Light is diffused when it reflects from a rough surface.

Conceptual Questions

Exercise:

Problem:

Using the law of reflection, explain how powder takes the shine off of a person’s nose. What is the name of the optical effect?

Solution:

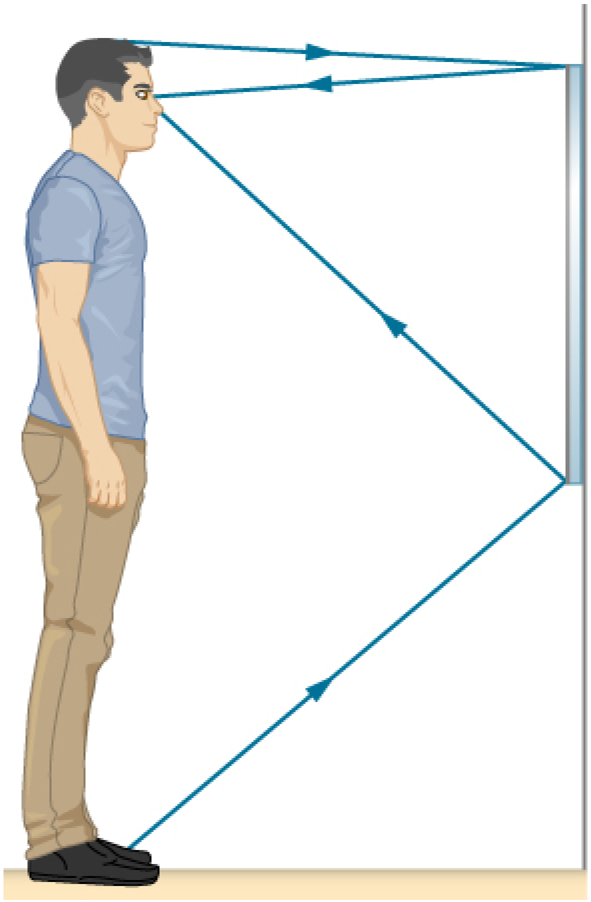
Powder consists of many small particles with randomly oriented surfaces. This leads to diffuse reflection, reducing shine.

Problems

Exercise:

Problem:

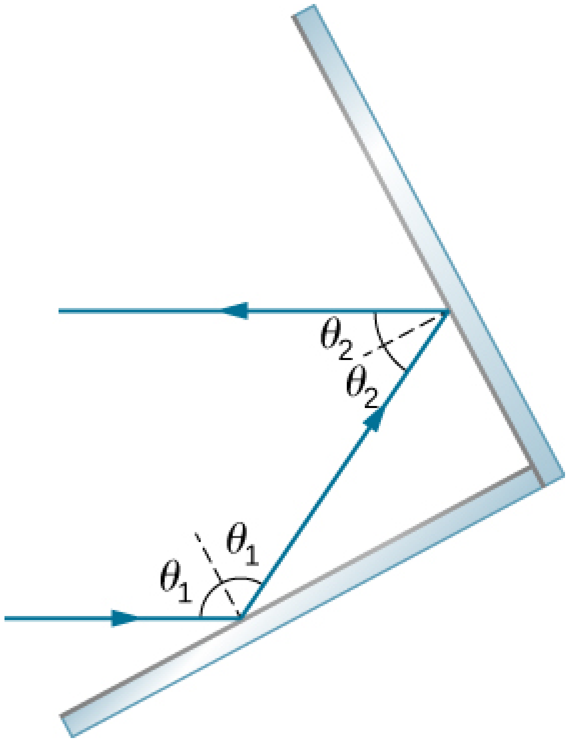
Suppose a man stands in front of a mirror as shown below. His eyes are 1.65 m above the floor and the top of his head is 0.13 m higher. Find the height above the floor of the top and bottom of the smallest mirror in which he can see both the top of his head and his feet. How is this distance related to the man’s height?



Exercise:

Problem:

Show that when light reflects from two mirrors that meet each other at a right angle, the outgoing ray is parallel to the incoming ray, as illustrated below.



Solution:

proof

Exercise:

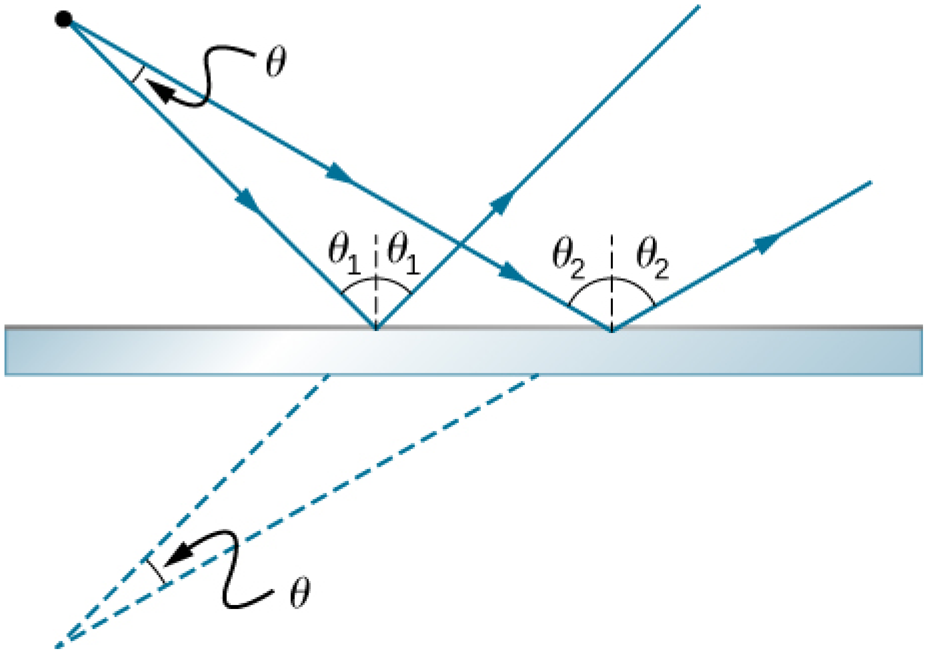
Problem:

On the Moon’s surface, lunar astronauts placed a corner reflector, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. What percent correction is needed to account for the delay in time due to the slowing of light in Earth’s atmosphere? Assume the distance to the Moon is precisely 3.84×10^8 m and Earth’s atmosphere (which varies in density with altitude) is equivalent to a layer 30.0 km thick with a constant index of refraction $n = 1.000293$.

Exercise:

Problem:

A flat mirror is neither converging nor diverging. To prove this, consider two rays originating from the same point and diverging at an angle θ (see below). Show that after striking a plane mirror, the angle between their directions remains θ.



Solution:

proof

Glossary

corner reflector
object consisting of two (or three) mutually perpendicular reflecting surfaces, so that the light that enters is reflected back exactly parallel to the direction from which it came

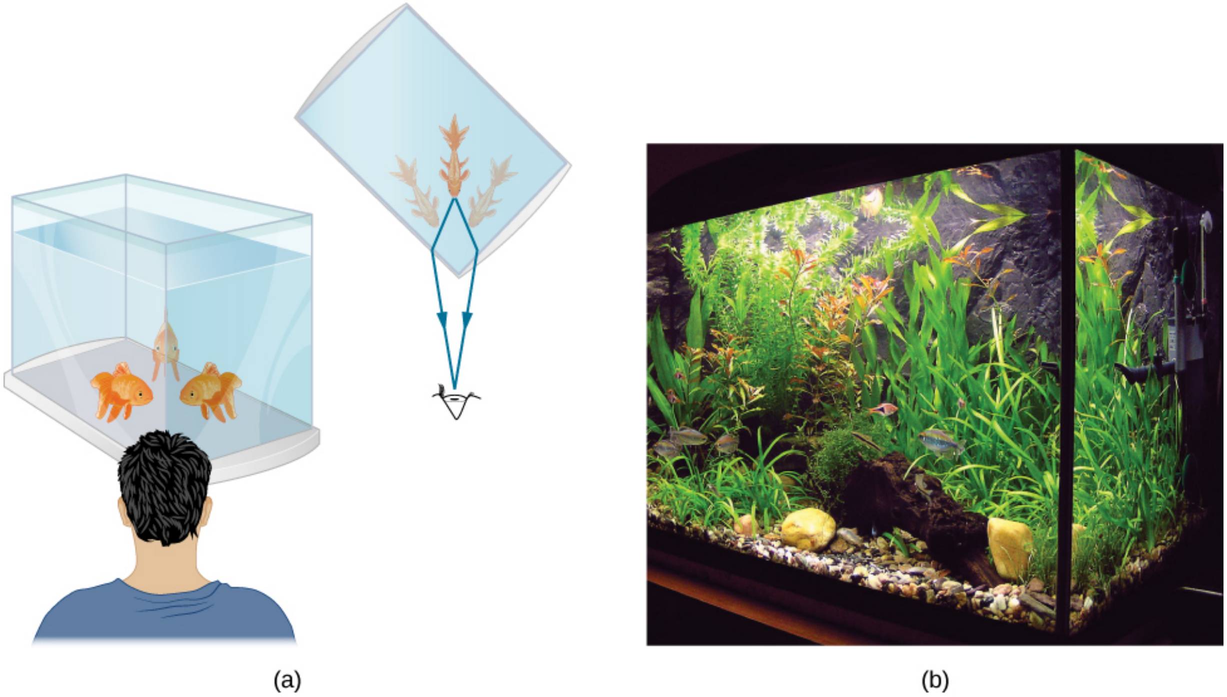
law of reflection
angle of reflection equals the angle of incidence

Refraction

By the end of this section, you will be able to:

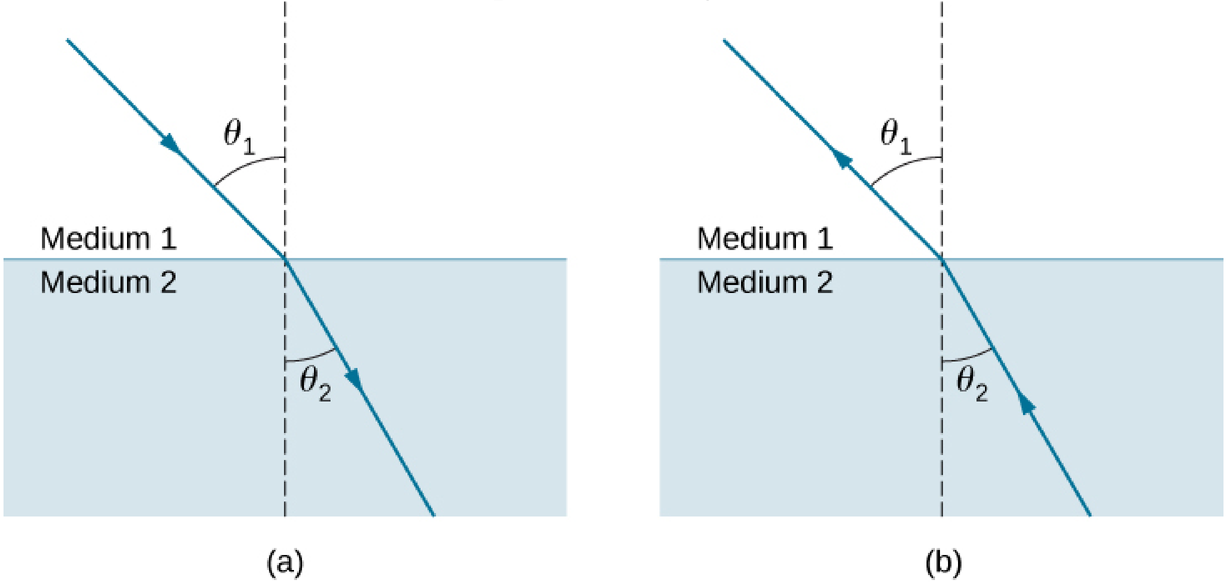
- Describe how rays change direction upon entering a medium
- Apply the law of refraction in problem solving

You may often notice some odd things when looking into a fish tank. For example, you may see the same fish appearing to be in two different places ([link]). This happens because light coming from the fish to you changes direction when it leaves the tank, and in this case, it can travel two different paths to get to your eyes. The changing of a light ray's direction (loosely called bending) when it passes through substances of different refractive indices is called **refraction** and is related to changes in the speed of light, $v = c/n$. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to data transmission through optical fibers.



(a) Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, the light can reach the observer by two different paths, so the fish seems to be in two different places. This bending of light is called refraction and is responsible for many optical phenomena. (b) This image shows refraction of light from a fish near the top of a fish tank.

[link] shows how a ray of light changes direction when it passes from one medium to another. As before, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it. (Some of the incident light is reflected from the surface, but for now we concentrate on the light that is transmitted.) The change in direction of the light ray depends on the relative values of the indices of refraction (The Propagation of Light) of the two media involved. In the situations shown, medium 2 has a greater index of refraction than medium 1. Note that as shown in [link](a), the direction of the ray moves closer to the perpendicular when it progresses from a medium with a lower index of refraction to one with a higher index of refraction. Conversely, as shown in [link](b), the direction of the ray moves away from the perpendicular when it progresses from a medium with a higher index of refraction to one with a lower index of refraction. The path is exactly reversible.



The change in direction of a light ray depends on how the index of refraction changes when it crosses from one medium to another. In the situations shown here, the index of refraction is greater in medium 2 than in medium 1. (a) A ray of light moves closer to the perpendicular when entering a medium with a higher index of refraction. (b) A ray of light moves away from the perpendicular when entering a medium with a lower index of refraction.

The amount that a light ray changes its direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction and thus a large change in angle. The exact mathematical relationship is the **law of refraction**, or Snell's law, after the Dutch mathematician Willebrord Snell (1591–1626), who discovered it in 1621. The law of refraction is stated in equation form as

Note:
Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Here n_1 and n_2 are the indices of refraction for media 1 and 2, and θ_1 and θ_2 are the angles between the rays and the perpendicular in media 1 and 2. The incoming ray is called the incident ray, the outgoing ray is called the refracted ray, and the associated angles are the incident angle and the refracted angle, respectively.

Snell's experiments showed that the law of refraction is obeyed and that a characteristic index of refraction *n* could be assigned to a given medium and its value measured. Snell was not aware that the speed of light varied in different media, a key fact used when we derive the law of refraction theoretically using Huygens's principle in Huygens's Principle.

Example:

Determining the Index of Refraction

Find the index of refraction for medium 2 in [link](a), assuming medium 1 is air and given that the incident angle is 30.0° and the angle of refraction is 22.0°.

Strategy

The index of refraction for air is taken to be 1 in most cases (and up to four significant figures, it is 1.000). Thus, $n_1 = 1.00$ here. From the given information, $\theta_1 = 30.0^\circ$ and $\theta_2 = 22.0^\circ$. With this information, the only unknown in Snell's law is n_2, so we can use Snell's law to find it.

Solution

From Snell's law we have

Equation:

$$\begin{array}{rcl} n_1 \sin\theta_1 & = & n_2 \sin\theta_2 \\ n_2 & = & n_1 \frac{\sin\theta_1}{\sin\theta_2}. \end{array}$$

Entering known values,

Equation:

$$n_2 = 1.00 \frac{\sin 30.0^\circ}{\sin 22.0^\circ} = \frac{0.500}{0.375} = 1.33.$$

Significance

This is the index of refraction for water, and Snell could have determined it by measuring the angles and performing this calculation. He would then have found 1.33 to be the appropriate index of refraction for water in all other situations, such as when a ray passes from water to glass. Today, we can verify that the index of refraction is related to the speed of light in a medium by measuring that speed directly.

Note:

Explore bending of light between two media with different indices of refraction. Use the "Intro" simulation and see how changing from air to water to glass changes the bending angle. Use the protractor tool to measure the angles and see if you can recreate the configuration in [link]. Also by measurement, confirm that the angle of reflection equals the angle of incidence.

Example:

A Larger Change in Direction

Suppose that in a situation like that in [link], light goes from air to diamond and that the incident angle is 30.0°. Calculate the angle of refraction θ_2 in the diamond.

Strategy

Again, the index of refraction for air is taken to be $n_1 = 1.00$, and we are given $\theta_1 = 30.0^\circ$. We can look up the index of refraction for diamond in [link], finding $n_2 = 2.419$. The only unknown in Snell's law is θ_2, which we wish to determine.

Solution

Solving Snell's law for $\sin\theta_2$ yields

Equation:

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1.$$

Entering known values,

Equation:

$$\sin\theta_2 = \frac{1.00}{2.419}\sin 30.0^\circ = (0.413)(0.500) = 0.207.$$

The angle is thus

Equation:

$$\theta_2 = \sin^{-1}(0.207) = 11.9^\circ.$$

Significance

For the same 30.0° angle of incidence, the angle of refraction in diamond is significantly smaller than in water $(11.9^\circ$ rather than 22.0°—see [link]). This means there is a larger change in direction in diamond. The cause of a large change in direction is a large change in the index of refraction (or speed). In general, the larger the change in speed, the greater the effect on the direction of the ray.

Note:
Exercise:

Problem:

Check Your Understanding In [link], the solid with the next highest index of refraction after diamond is zircon. If the diamond in [link] were replaced with a piece of zircon, what would be the new angle of refraction?

Solution:

15.1°

Summary

- The change of a light ray’s direction when it passes through variations in matter is called refraction.
- The law of refraction, also called Snell’s law, relates the indices of refraction for two media at an interface to the change in angle of a light ray passing through that interface.

Conceptual Questions

Exercise:

Problem:

Diffusion by reflection from a rough surface is described in this chapter. Light can also be diffused by refraction. Describe how this occurs in a specific situation, such as light interacting with crushed ice.

Exercise:

Problem:

Will light change direction toward or away from the perpendicular when it goes from air to water? Water to glass? Glass to air?

Solution:

“toward” when increasing *n* (air to water, water to glass); “away” when decreasing *n* (glass to air)

Exercise:

Problem:

Explain why an object in water always appears to be at a depth shallower than it actually is?

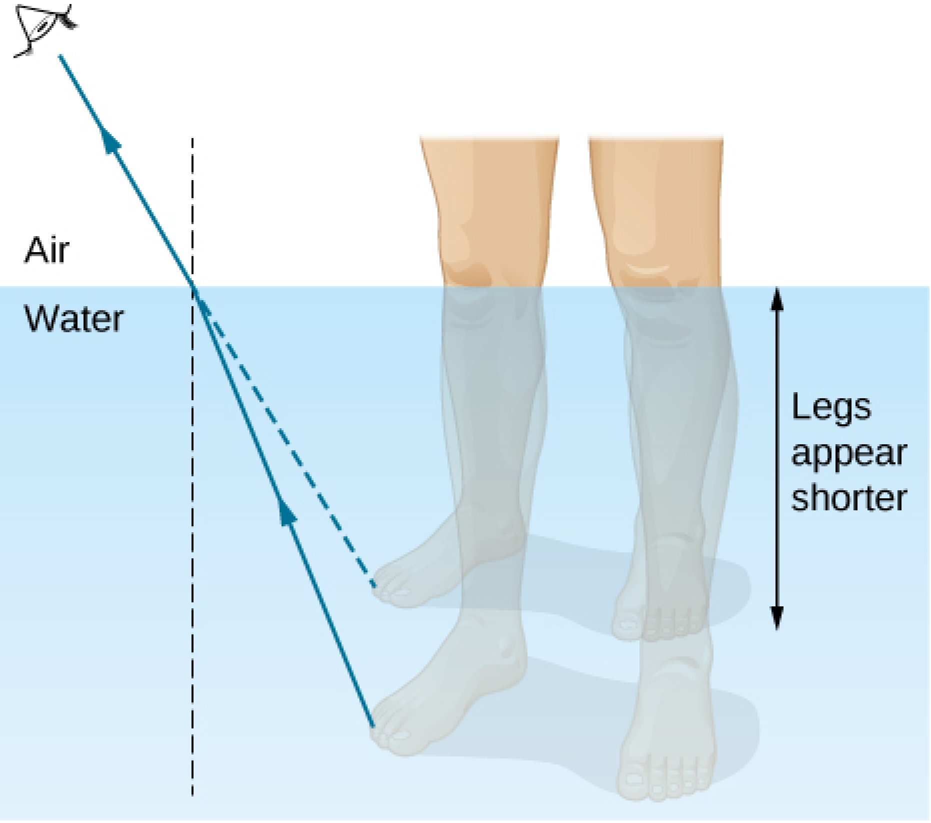
Exercise:

Problem:

Explain why a person’s legs appear very short when wading in a pool. Justify your explanation with a ray diagram showing the path of rays from the feet to the eye of an observer who is out of the water.

Solution:

A ray from a leg emerges from water after refraction. The observer in air perceives an apparent location for the source, as if a ray traveled in a straight line. See the dashed ray below.



Exercise:

Problem:

Explain why an oar that is partially submerged in water appears bent.

Problems

Unless otherwise specified, for problems 1 through 10, the indices of refraction of glass and water should be taken to be 1.50 and 1.333, respectively.

Exercise:

Problem:

A light beam in air has an angle of incidence of 35° at the surface of a glass plate. What are the angles of reflection and refraction?

Exercise:

Problem:

A light beam in air is incident on the surface of a pond, making an angle of 20° with respect to the surface. What are the angles of reflection and refraction?

Solution:

reflection, 70°; refraction, 45°

Exercise:

Problem:

When a light ray crosses from water into glass, it emerges at an angle of 30° with respect to the normal of the interface. What is its angle of incidence?

Exercise:

Problem:

A pencil flashlight submerged in water sends a light beam toward the surface at an angle of incidence of 30°. What is the angle of refraction in air?

Solution:

42°

Exercise:

Problem:

Light rays from the Sun make a 30° angle to the vertical when seen from below the surface of a body of water. At what angle above the horizon is the Sun?

Exercise:

Problem:

The path of a light beam in air goes from an angle of incidence of 35° to an angle of refraction of 22° when it enters a rectangular block of plastic. What is the index of refraction of the plastic?

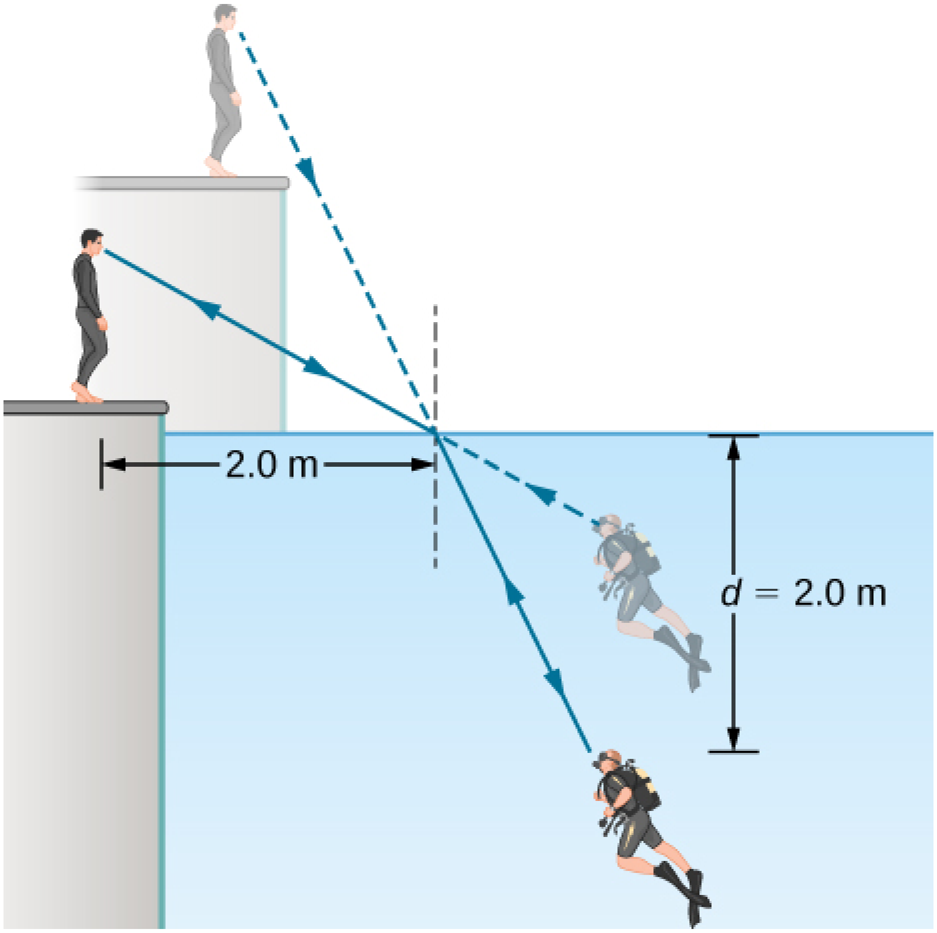
Solution:

1.53

Exercise:

Problem:

A scuba diver training in a pool looks at his instructor as shown below. What angle does the ray from the instructor’s face make with the perpendicular to the water at the point where the ray enters? The angle between the ray in the water and the perpendicular to the water is 25.0°.



Exercise:

Problem:

(a) Using information in the preceding problem, find the height of the instructor's head above the water, noting that you will first have to calculate the angle of incidence. (b) Find the apparent depth of the diver's head below water as seen by the instructor.

Solution:

a. 2.9 m; b. 1.4 m

Glossary

law of refraction

when a light ray crosses from one medium to another, it changes direction by an amount that depends on the index of refraction of each medium and the sines of the angle of incidence and angle of refraction

refraction
changing of a light ray’s direction when it passes through variations in matter

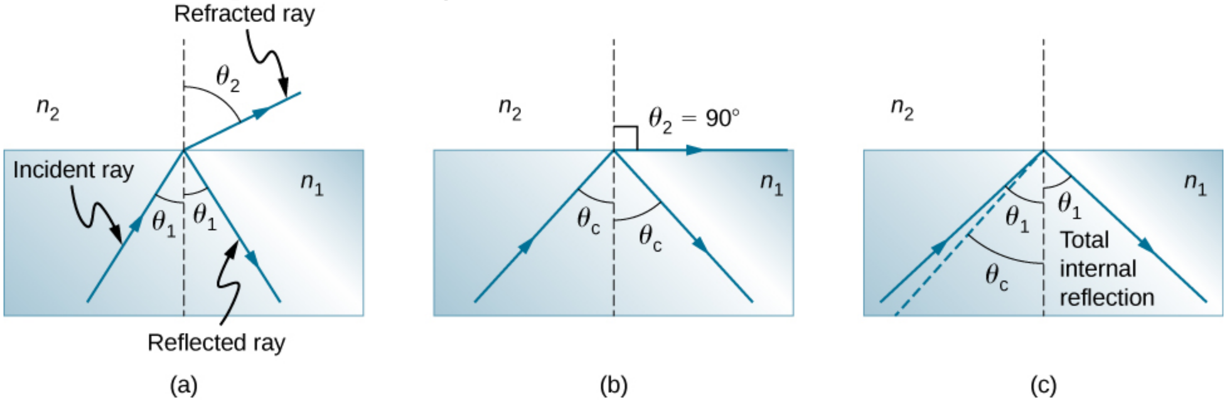
Total Internal Reflection

By the end of this section, you will be able to:

- Explain the phenomenon of total internal reflection
- Describe the workings and uses of optical fibers
- Analyze the reason for the sparkle of diamonds

A good-quality mirror may reflect more than 90% of the light that falls on it, absorbing the rest. But it would be useful to have a mirror that reflects all of the light that falls on it. Interestingly, we can produce total reflection using an aspect of refraction.

Consider what happens when a ray of light strikes the surface between two materials, as shown in [link](a). Part of the light crosses the boundary and is refracted; the rest is reflected. If, as shown in the figure, the index of refraction for the second medium is less than for the first, the ray bends away from the perpendicular. (Since $n_1 > n_2$, the angle of refraction is greater than the angle of incidence—that is, $\theta_2 > \theta_1$.) Now imagine what happens as the incident angle increases. This causes θ_2 to increase also. The largest the angle of refraction θ_2 can be is 90°, as shown in part (b). The **critical angle** θ_c for a combination of materials is defined to be the incident angle θ_1 that produces an angle of refraction of 90°. That is, θ_c is the incident angle for which $\theta_2 = 90^\circ$. If the incident angle θ_1 is greater than the critical angle, as shown in [link](c), then all of the light is reflected back into medium 1, a condition called **total internal reflection**. (As the figure shows, the reflected rays obey the law of reflection so that the angle of reflection is equal to the angle of incidence in all three cases.)



(a) A ray of light crosses a boundary where the index of refraction decreases. That is, $n_2 < n_1$. The ray bends away from the perpendicular. (b) The critical angle θ_{c} is the angle of incidence for which the angle of refraction is 90°. (c) Total internal reflection occurs when the incident angle is greater than the critical angle.

Snell's law states the relationship between angles and indices of refraction. It is given by

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

When the incident angle equals the critical angle $(\theta_1 = \theta_c)$, the angle of refraction is 90° $(\theta_2 = 90^\circ)$. Noting that $\sin 90^\circ = 1$, Snell's law in this case becomes

Equation:

$$n_1 \sin \theta_1 = n_2.$$

The critical angle θ_{c} for a given combination of materials is thus

Note:

Equation:

$$\theta_{\text{c}} = \sin^{-1}\left(\frac{n_2}{n_1}\right) \text{ for } n_1 > n_2.$$

Total internal reflection occurs for any incident angle greater than the critical angle θ_{c}, and it can only occur when the second medium has an index of refraction less than the first. Note that this equation is written for a light ray that travels in medium 1 and reflects from medium 2, as shown in [link].

Example:

Determining a Critical Angle

What is the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air? The index of refraction for polystyrene is 1.49.

Strategy

The index of refraction of air can be taken to be 1.00, as before. Thus, the condition that the second medium (air) has an index of refraction less than the first (plastic) is satisfied, and we can use the equation

Equation:

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right)$$

to find the critical angle θ_c, where $n_2 = 1.00$ and $n_1 = 1.49$.

Solution

Substituting the identified values gives

Equation:

$$\theta_c = \sin^{-1}\left(\frac{1.00}{1.49}\right) = \sin^{-1}(0.671) = 42.2^\circ.$$

Significance

This result means that any ray of light inside the plastic that strikes the surface at an angle greater than 42.2° is totally reflected. This makes the inside surface of the clear plastic a perfect mirror for such rays, without any need for the silvering used on common mirrors. Different combinations of materials have different critical angles, but any combination with $n_1 > n_2$ can produce total internal reflection. The same calculation as made here shows that the critical angle for a ray going from water to air is 48.6°, whereas that from diamond to air is 24.4°, and that from flint glass to crown glass is 66.3°.

Note:
Exercise:

Problem:

Check Your Understanding At the surface between air and water, light rays can go from air to water and from water to air. For which ray is there no possibility of total internal reflection?

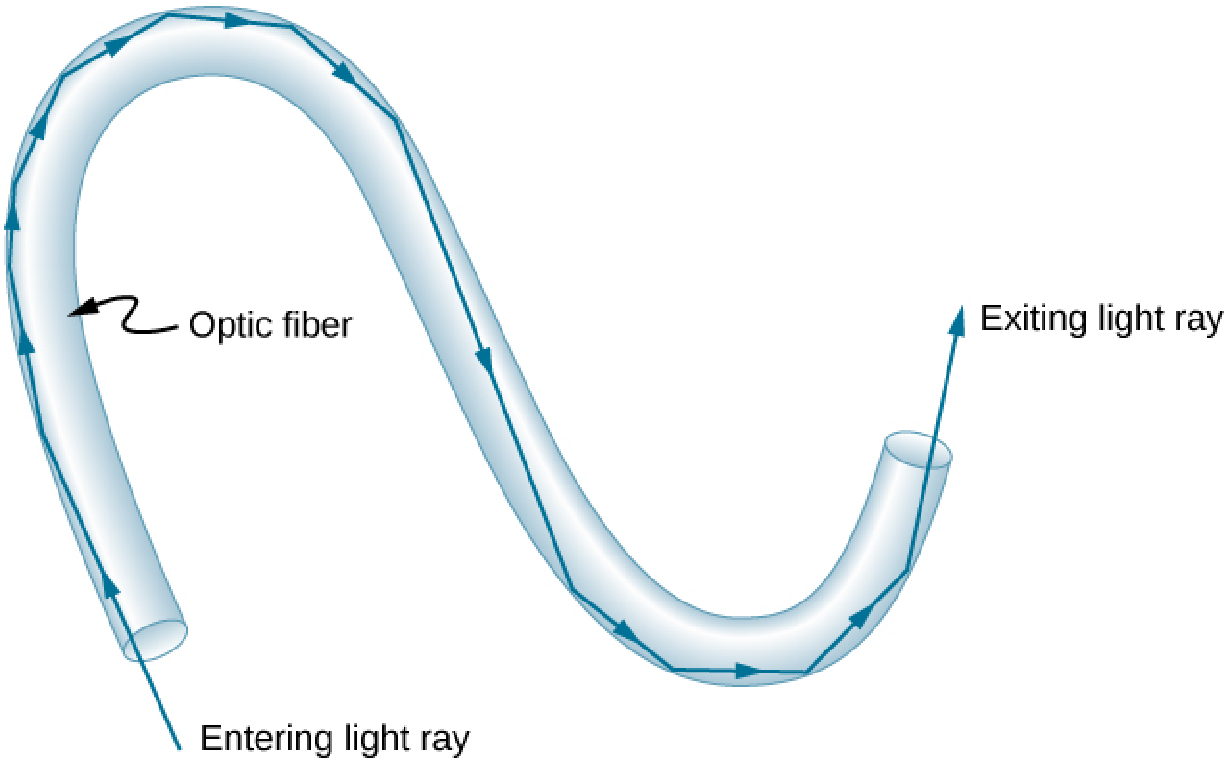
Solution:

air to water, because the condition that the second medium must have a smaller index of refraction is not satisfied

In the photo that opens this chapter, the image of a swimmer underwater is captured by a camera that is also underwater. The swimmer in the upper half of the photograph, apparently facing upward, is, in fact, a reflected image of the swimmer below. The circular ripple near the photograph's center is actually on the water surface. The undisturbed water surrounding it makes a good reflecting surface when viewed from below, thanks to total internal reflection. However, at the very top edge of this photograph, rays from below strike the surface with incident angles less than the critical angle, allowing the camera to capture a view of activities on the pool deck above water.

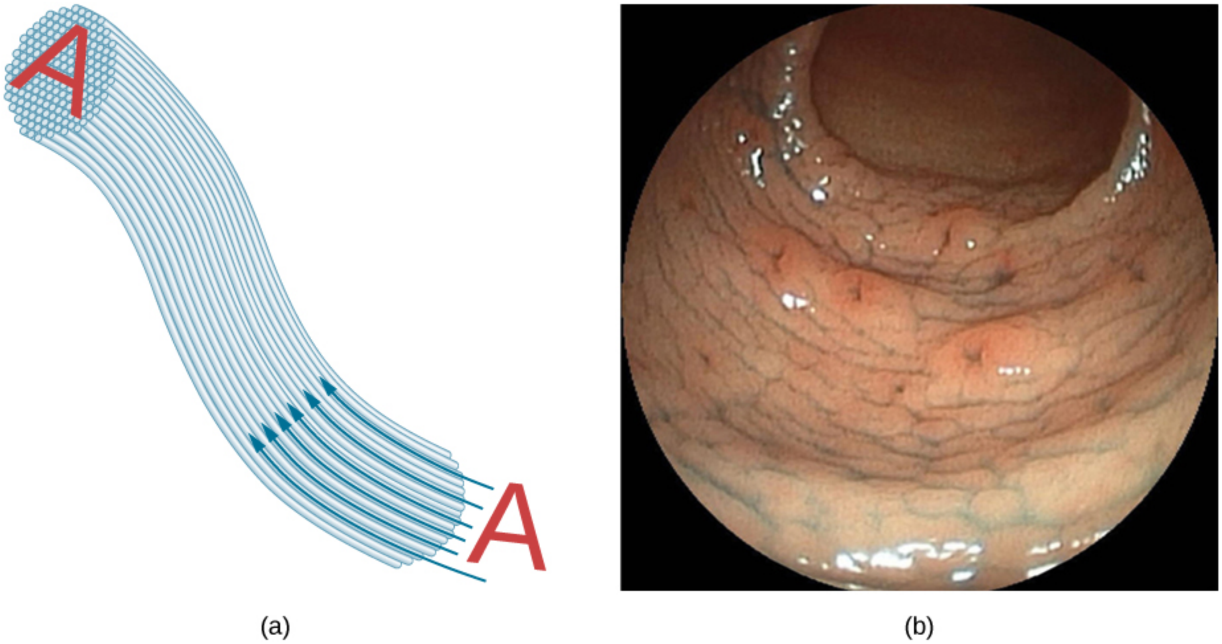
Fiber Optics: Endoscopes to Telephones

Fiber optics is one application of total internal reflection that is in wide use. In communications, it is used to transmit telephone, internet, and cable TV signals. **Fiber optics** employs the transmission of light down fibers of plastic or glass. Because the fibers are thin, light entering one is likely to strike the inside surface at an angle greater than the critical angle and, thus, be totally reflected ([link]). The index of refraction outside the fiber must be smaller than inside. In fact, most fibers have a varying refractive index to allow more light to be guided along the fiber through total internal refraction. Rays are reflected around corners as shown, making the fibers into tiny light pipes.



Light entering a thin optic fiber may strike the inside surface at large or grazing angles and is completely reflected if these angles exceed the critical angle. Such rays continue down the fiber, even following it around corners, since the angles of reflection and incidence remain large.

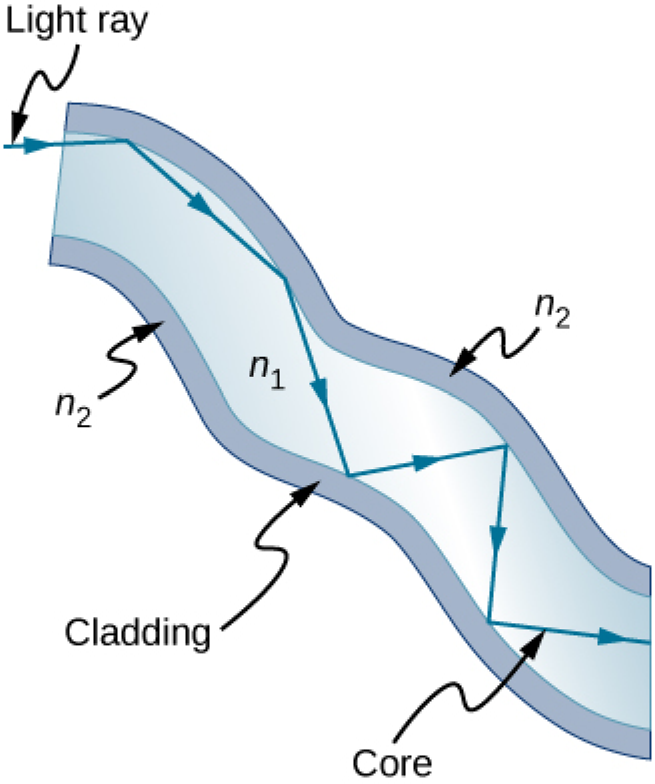
Bundles of fibers can be used to transmit an image without a lens, as illustrated in [link]. The output of a device called an endoscope is shown in [link](b). Endoscopes are used to explore the interior of the body through its natural orifices or minor incisions. Light is transmitted down one fiber bundle to illuminate internal parts, and the reflected light is transmitted back out through another bundle to be observed.



(a) An image “A” is transmitted by a bundle of optical fibers. (b) An endoscope is used to probe the body, both transmitting light to the interior and returning an image such as the one shown of a human epiglottis (a structure at the base of the tongue). (credit b: modification of work by “Med_Chaos”/Wikimedia Commons)

Fiber optics has revolutionized surgical techniques and observations within the body, with a host of medical diagnostic and therapeutic uses. Surgery can be performed, such as arthroscopic surgery on a knee or shoulder joint, employing cutting tools attached to and observed with the endoscope. Samples can also be obtained, such as by lassoing an intestinal polyp for external examination. The flexibility of the fiber optic bundle allows doctors to navigate it around small and difficult-to-reach regions in the body, such as the intestines, the heart, blood vessels, and joints. Transmission of an intense laser beam to burn away obstructing plaques in major arteries, as well as delivering light to activate chemotherapy drugs, are becoming commonplace. Optical fibers have in fact enabled microsurgery and remote surgery where the incisions are small and the surgeon’s fingers do not need to touch the diseased tissue.

Optical fibers in bundles are surrounded by a cladding material that has a lower index of refraction than the core ([link]). The cladding prevents light from being transmitted between fibers in a bundle. Without cladding, light could pass between fibers in contact, since their indices of refraction are identical. Since no light gets into the cladding (there is total internal reflection back into the core), none can be transmitted between clad fibers that are in contact with one another. Instead, the light is propagated along the length of the fiber, minimizing the loss of signal and ensuring that a quality image is formed at the other end. The cladding and an additional protective layer make optical fibers durable as well as flexible.



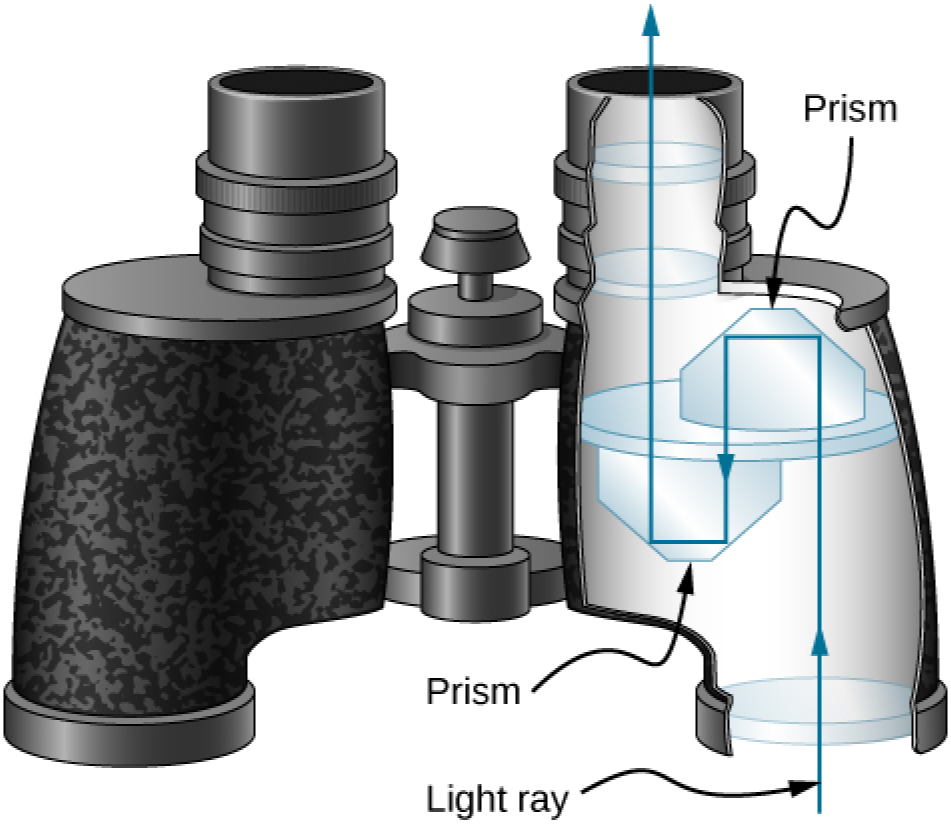
Fibers in bundles are clad by a material that has a lower index of refraction than the core to ensure total internal reflection, even when fibers are in contact with one another.

Special tiny lenses that can be attached to the ends of bundles of fibers have been designed and fabricated. Light emerging from a fiber bundle can be focused through such a lens, imaging a tiny spot. In some cases, the spot can be scanned, allowing quality imaging of a region inside the body. Special minute optical filters inserted at the end of the fiber bundle have the capacity to image the interior of organs located tens of microns below the surface without cutting the surface—an area known as nonintrusive diagnostics. This is particularly useful for determining the extent of cancers in the stomach and bowel.

In another type of application, optical fibers are commonly used to carry signals for telephone conversations and internet communications. Extensive optical fiber cables have been placed on the ocean floor and underground to enable optical communications. Optical fiber communication systems offer several advantages over electrical (copper)-based systems, particularly for long distances. The fibers can be made so transparent that light can travel many kilometers before it becomes dim enough to require amplification—much superior to copper conductors. This property of optical fibers is called low loss. Lasers emit light with characteristics that allow far more conversations in one fiber than are possible with electric signals on a single conductor. This property of optical fibers is called high bandwidth. Optical signals in one fiber do not produce undesirable effects in other adjacent fibers. This property of optical fibers is called reduced crosstalk. We shall explore the unique characteristics of laser radiation in a later chapter.

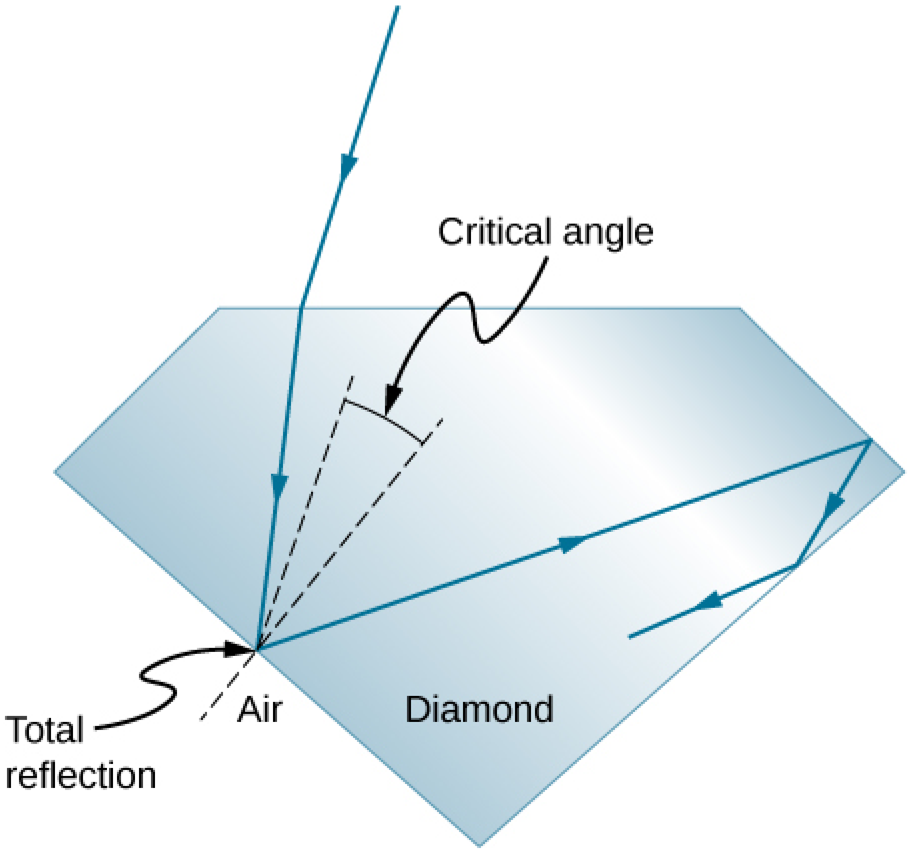
Corner Reflectors and Diamonds

Corner reflectors (The Law of Reflection) are perfectly efficient when the conditions for total internal reflection are satisfied. With common materials, it is easy to obtain a critical angle that is less than 45°. One use of these perfect mirrors is in binoculars, as shown in [link]. Another use is in periscopes found in submarines.



These binoculars employ corner reflectors (prisms) with total internal reflection to get light to the observer’s eyes.

Total internal reflection, coupled with a large index of refraction, explains why diamonds sparkle more than other materials. The critical angle for a diamond-to-air surface is only 24.4°, so when light enters a diamond, it has trouble getting back out ([link]). Although light freely enters the diamond, it can exit only if it makes an angle less than 24.4°. Facets on diamonds are specifically intended to make this unlikely. Good diamonds are very clear, so that the light makes many internal reflections and is concentrated before exiting—hence the bright sparkle. (Zircon is a natural gemstone that has an exceptionally large index of refraction, but it is not as large as diamond, so it is not as highly prized. Cubic zirconia is manufactured and has an even higher index of refraction (≈ 2.17), but it is still less than that of diamond.) The colors you see emerging from a clear diamond are not due to the diamond's color, which is usually nearly colorless. The colors result from dispersion, which we discuss in Dispersion. Colored diamonds get their color from structural defects of the crystal lattice and the inclusion of minute quantities of graphite and other materials. The Argyle Mine in Western Australia produces around 90% of the world's pink, red, champagne, and cognac diamonds, whereas around 50% of the world's clear diamonds come from central and southern Africa.



Light cannot easily escape a diamond, because its critical angle with air is so small. Most reflections are total, and the facets are placed so that light can exit only in particular ways—thus concentrating the light and making the diamond sparkle brightly.

Note:
Explore refraction and reflection of light between two media with different indices of refraction. Try to make the refracted ray disappear with total internal reflection. Use the protractor tool to measure the critical angle and compare with the prediction from [link].

Summary

- The incident angle that produces an angle of refraction of 90° is called the critical angle.
- Total internal reflection is a phenomenon that occurs at the boundary between two media, such that if the incident angle in the first medium is greater than the critical angle, then all the light is reflected back into that medium.
- Fiber optics involves the transmission of light down fibers of plastic or glass, applying the principle of total internal reflection.
- Cladding prevents light from being transmitted between fibers in a bundle.
- Diamonds sparkle due to total internal reflection coupled with a large index of refraction.

Conceptual Questions

Exercise:

Problem:

A ring with a colorless gemstone is dropped into water. The gemstone becomes invisible when submerged. Can it be a diamond? Explain.

Solution:

The gemstone becomes invisible when its index of refraction is the same, or at least similar to, the water surrounding it. Because diamond has a particularly high index of refraction, it can still sparkle as a result of total internal reflection, not invisible.

Exercise:

Problem:

The most common type of mirage is an illusion that light from faraway objects is reflected by a pool of water that is not really there. Mirages are generally observed in deserts, when there is a hot layer of air near the ground. Given that the refractive index of air is lower for air at higher temperatures, explain how mirages can be formed.

Exercise:

Problem:

How can you use total internal reflection to estimate the index of refraction of a medium?

Solution:

One can measure the critical angle by looking for the onset of total internal reflection as the angle of incidence is varied. [link] can then be applied to compute the index of refraction.

Problems

Exercise:

Problem:

Verify that the critical angle for light going from water to air is 48.6°, as discussed at the end of [link], regarding the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air.

Exercise:

Problem:

(a) At the end of [link], it was stated that the critical angle for light going from diamond to air is 24.4°. Verify this. (b) What is the critical angle for light going from zircon to air?

Solution:

a. 24.42°; b. 31.33°

Exercise:

Problem:

An optical fiber uses flint glass clad with crown glass. What is the critical angle?

Exercise:

Problem:

At what minimum angle will you get total internal reflection of light traveling in water and reflected from ice?

Solution:

79.11°

Exercise:

Problem:

Suppose you are using total internal reflection to make an efficient corner reflector. If there is air outside and the incident angle is 45.0°, what must be the minimum index of refraction of the material from which the reflector is made?

Exercise:

Problem:

You can determine the index of refraction of a substance by determining its critical angle. (a) What is the index of refraction of a substance that has a critical angle of 68.4° when submerged in water? What is the substance, based on [link]? (b) What would the critical angle be for this substance in air?

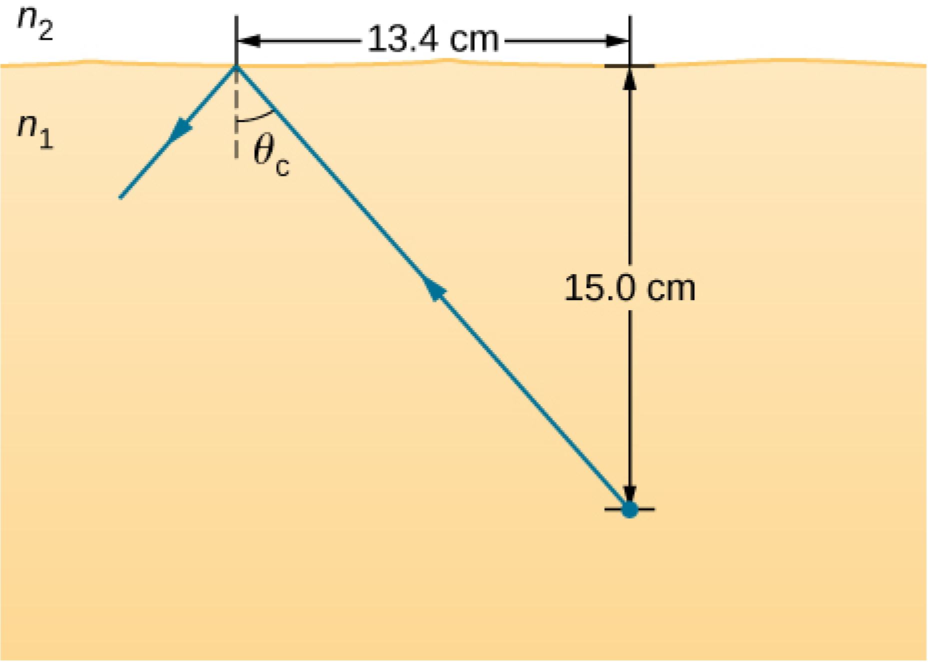
Solution:

a. 1.43, fluorite; b. 44.2°

Exercise:

Problem:

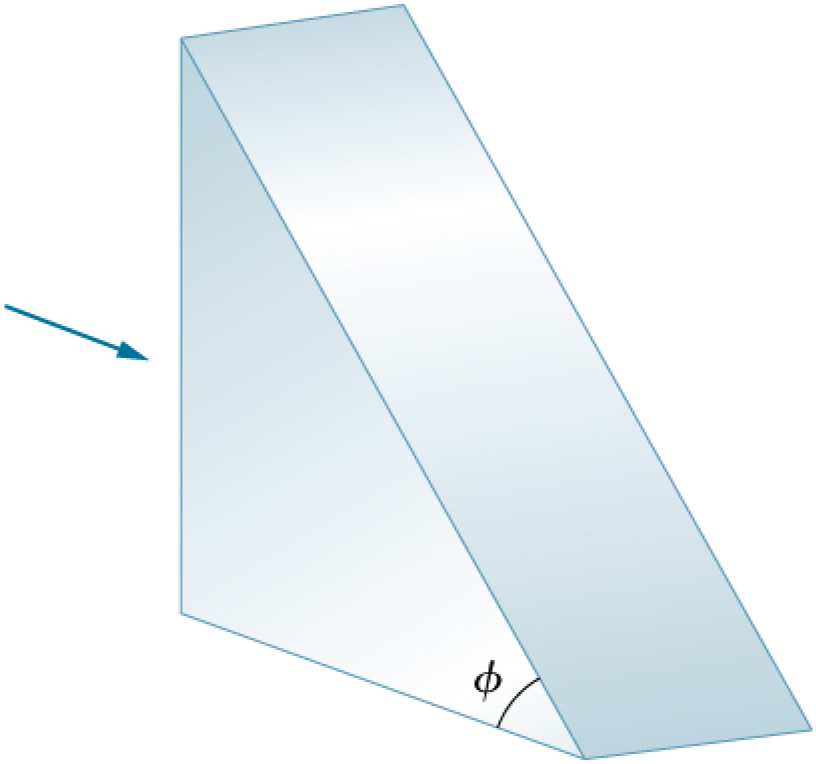
A ray of light, emitted beneath the surface of an unknown liquid with air above it, undergoes total internal reflection as shown below. What is the index of refraction for the liquid and its likely identification?



Exercise:

Problem:

Light rays fall normally on the vertical surface of the glass prism $(n = 1.50)$ shown below. (a) What is the largest value for ϕ such that the ray is totally reflected at the slanted face? (b) Repeat the calculation of part (a) if the prism is immersed in water.



Solution:

a. 48.2°; b. 27.3°

Glossary

critical angle
: incident angle that produces an angle of refraction of 90°

fiber optics
: field of study of the transmission of light down fibers of plastic or glass, applying the principle of total internal reflection

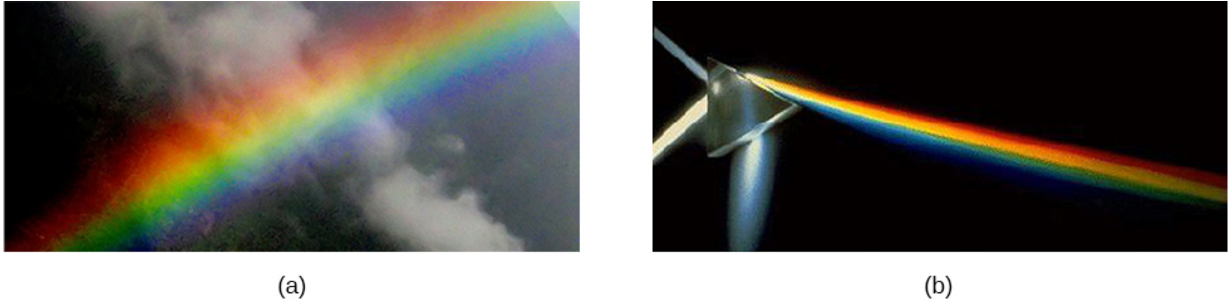
total internal reflection
: phenomenon at the boundary between two media such that all the light is reflected and no refraction occurs

Dispersion

By the end of this section, you will be able to:

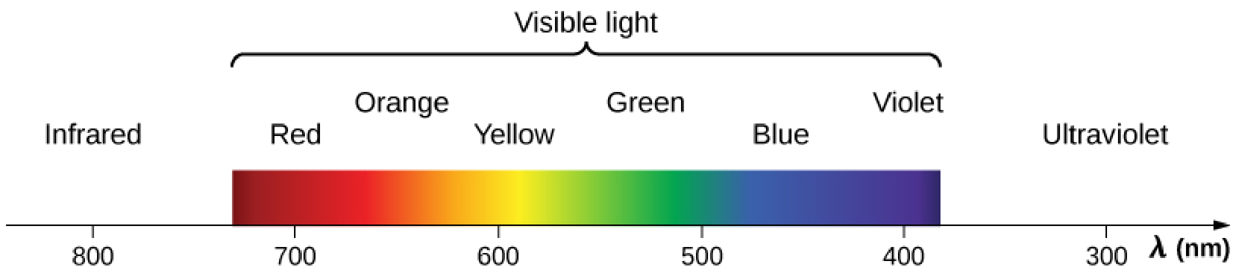
- Explain the cause of dispersion in a prism
- Describe the effects of dispersion in producing rainbows
- Summarize the advantages and disadvantages of dispersion

Everyone enjoys the spectacle of a rainbow glimmering against a dark stormy sky. How does sunlight falling on clear drops of rain get broken into the rainbow of colors we see? The same process causes white light to be broken into colors by a clear glass prism or a diamond ([link]).



The colors of the rainbow (a) and those produced by a prism (b) are identical. (credit a: modification of work by “Alfredo55”/Wikimedia Commons; credit b: modification of work by NASA)

We see about six colors in a rainbow—red, orange, yellow, green, blue, and violet; sometimes indigo is listed, too. These colors are associated with different wavelengths of light, as shown in [link]. When our eye receives pure-wavelength light, we tend to see only one of the six colors, depending on wavelength. The thousands of other hues we can sense in other situations are our eye’s response to various mixtures of wavelengths. White light, in particular, is a fairly uniform mixture of all visible wavelengths. Sunlight, considered to be white, actually appears to be a bit yellow, because of its mixture of wavelengths, but it does contain all visible wavelengths. The sequence of colors in rainbows is the same sequence as the colors shown in the figure. This implies that white light is spread out in a rainbow according to wavelength. **Dispersion** is defined as the spreading of white light into its full spectrum of wavelengths. More technically, dispersion occurs whenever the propagation of light depends on wavelength.



Even though rainbows are associated with six colors, the rainbow is a continuous distribution of colors according to wavelengths.

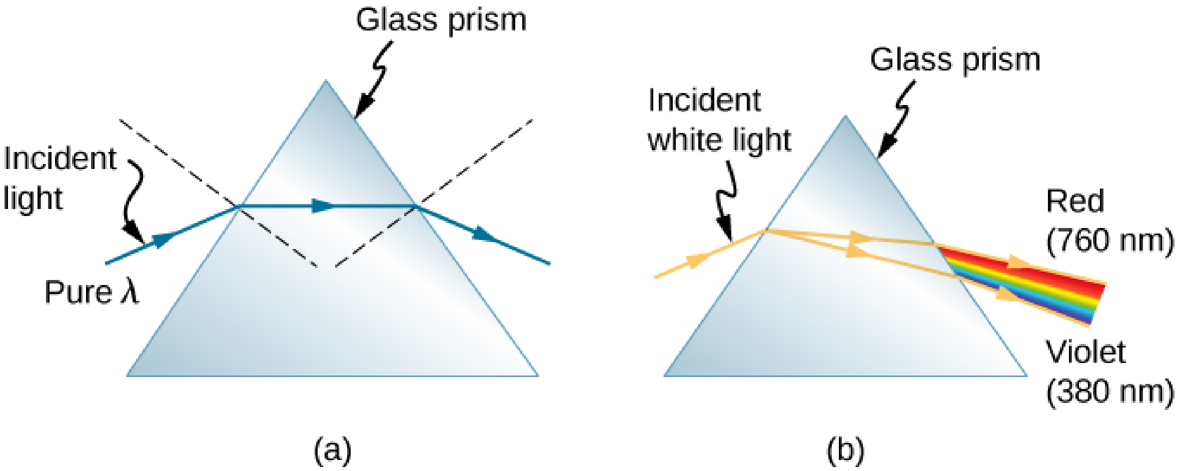
Any type of wave can exhibit dispersion. For example, sound waves, all types of electromagnetic waves, and water waves can be dispersed according to wavelength. Dispersion may require special circumstances and can result in spectacular displays such as in the production of a rainbow. This is also true for sound, since all frequencies ordinarily travel at the same speed. If you listen to sound through a long tube, such as a vacuum cleaner hose, you can easily hear it dispersed by interaction with the tube. Dispersion, in fact, can reveal a great deal about what the wave has encountered that disperses its wavelengths. The dispersion of electromagnetic radiation from outer space, for example, has revealed much about what exists between the stars—the so-called interstellar medium.

Note:
Nick Moore’s video discusses dispersion of a pulse as he taps a long spring. Follow his explanation as Moore replays the high-speed footage showing high frequency waves outrunning the lower frequency waves.

Refraction is responsible for dispersion in rainbows and many other situations. The angle of refraction depends on the index of refraction, as we know from Snell’s law. We know that the index of refraction n depends on the medium. But for a given medium, n also depends on wavelength ([link]). Note that for a given medium, n increases as wavelength decreases and is greatest for violet light. Thus, violet light is bent more than red light, as shown for a prism in [link](b). White light is dispersed into the same sequence of wavelengths as seen in [link] and [link].

Medium	**Red (660 nm)**	**Orange (610 nm)**	**Yellow (580 nm)**	**Green (550 nm)**	**Blue (470 nm)**	**Violet (410 nm)**
Water	1.331	1.332	1.333	1.335	1.338	1.342
Diamond	2.410	2.415	2.417	2.426	2.444	2.458
Glass, crown	1.512	1.514	1.518	1.519	1.524	1.530
Glass, flint	1.662	1.665	1.667	1.674	1.684	1.698
Polystyrene	1.488	1.490	1.492	1.493	1.499	1.506
Quartz, fused	1.455	1.456	1.458	1.459	1.462	1.468

Index of Refraction *n* in Selected Media at Various Wavelengths

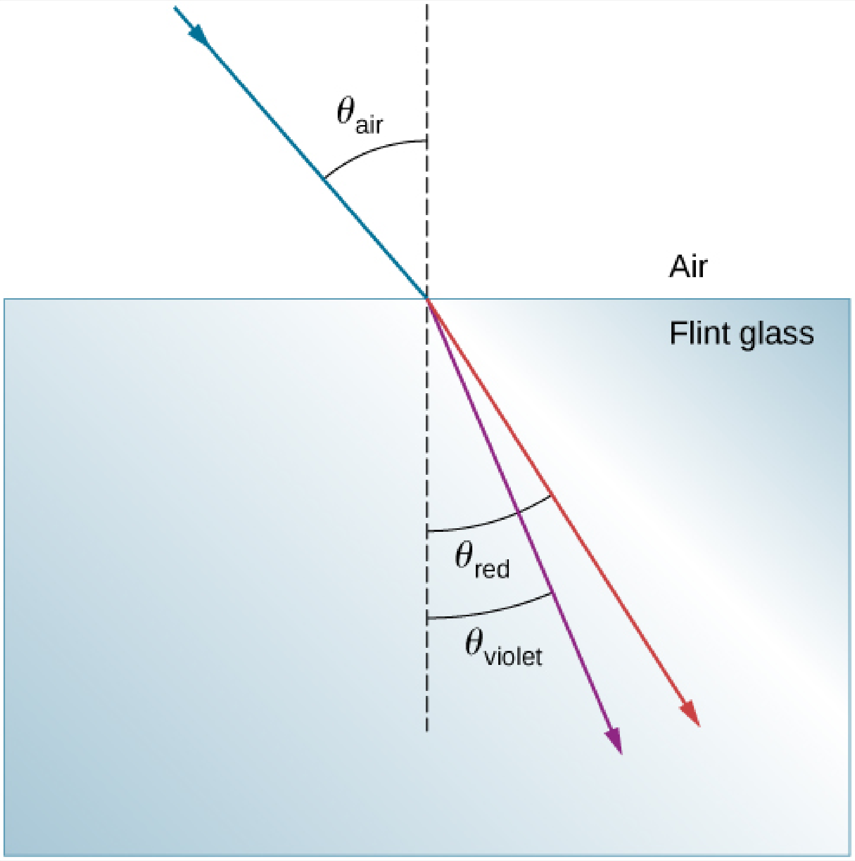


(a) A pure wavelength of light falls onto a prism and is refracted at both surfaces. (b) White light is dispersed by the prism (shown exaggerated). Since the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Example:

Dispersion of White Light by Flint Glass

A beam of white light goes from air into flint glass at an incidence angle of 43.2°. What is the angle between the red (660 nm) and violet (410 nm) parts of the refracted light?



Strategy

Values for the indices of refraction for flint glass at various wavelengths are listed in [link]. Use these values for calculate the angle of refraction for each color and then take the difference to find the dispersion angle.

Solution

Applying the law of refraction for the red part of the beam

Equation:

$$n_{\text{air}} \sin \theta_{\text{air}} = n_{\text{red}} \sin \theta_{\text{red}},$$

we can solve for the angle of refraction as

Equation:

$$\theta_{\text{red}} = \sin^{-1}\left(\frac{n_{\text{air}} \sin \theta_{\text{air}}}{n_{\text{red}}}\right) = \sin^{-1}\left[\frac{(1.000) \sin 43.2^\circ}{(1.662)}\right] = 27.0^\circ.$$

Similarly, the angle of incidence for the violet part of the beam is

Equation:

$$\theta_{\text{violet}} = \sin^{-1}\left(\frac{n_{\text{air}} \sin \theta_{\text{air}}}{n_{\text{violet}}}\right) = \sin^{-1}\left[\frac{(1.000) \sin 43.2^\circ}{(1.698)}\right] = 26.4^\circ.$$

The difference between these two angles is

Equation:

$$\theta_{\text{red}} - \theta_{\text{violet}} = 27.0^\circ - 26.4^\circ = 0.6^\circ.$$

Significance

Although 0.6° may seem like a negligibly small angle, if this beam is allowed to propagate a long enough distance, the dispersion of colors becomes quite noticeable.

Note:

Exercise:

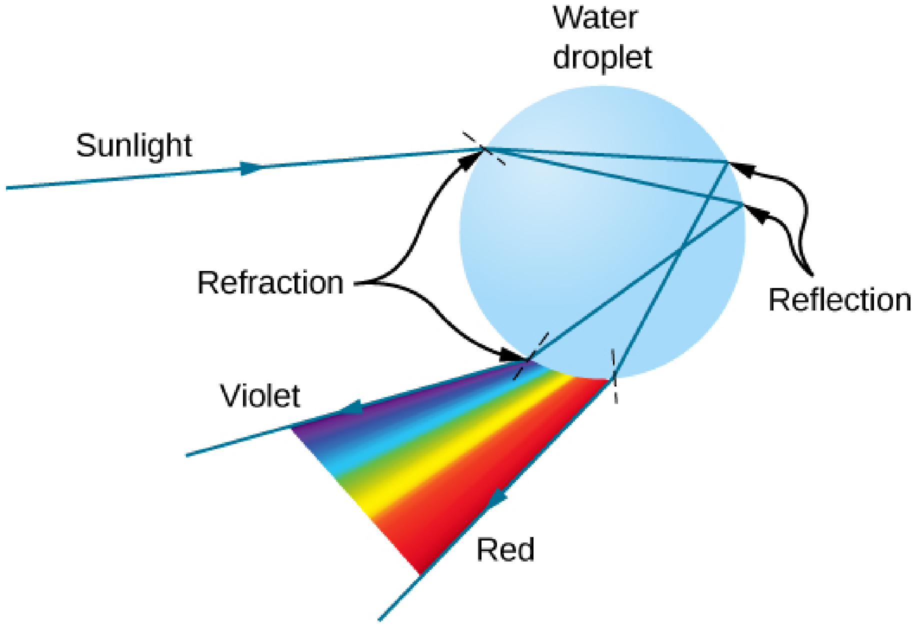
Problem:

Check Your Understanding In the preceding example, how much distance inside the block of flint glass would the red and the violet rays have to progress before they are separated by 1.0 mm?

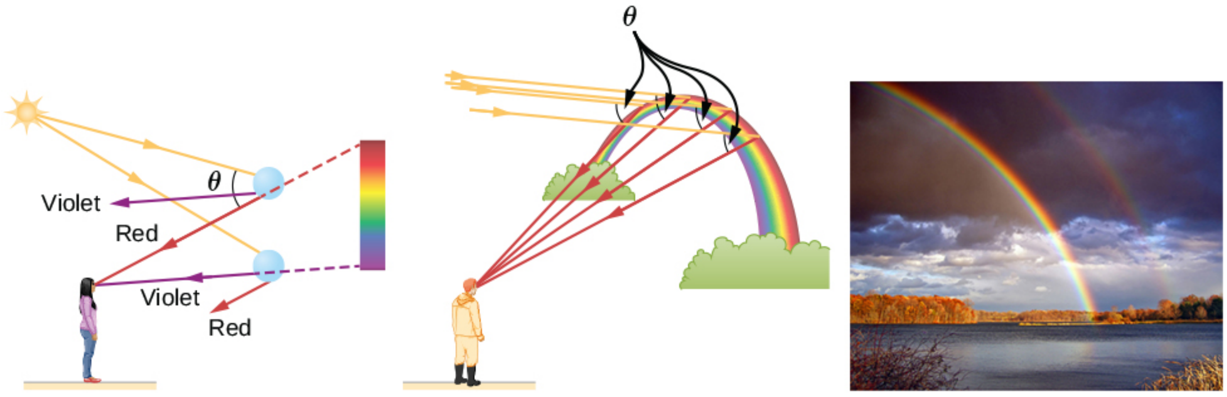
Solution:

9.3 cm

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you look away from the Sun. Light enters a drop of water and is reflected from the back of the drop ([link]). The light is refracted both as it enters and as it leaves the drop. Since the index of refraction of water varies with wavelength, the light is dispersed, and a rainbow is observed ([link](a)). (No dispersion occurs at the back surface, because the law of reflection does not depend on wavelength.) The actual rainbow of colors seen by an observer depends on the myriad rays being refracted and reflected toward the observer’s eyes from numerous drops of water. The effect is most spectacular when the background is dark, as in stormy weather, but can also be observed in waterfalls and lawn sprinklers. The arc of a rainbow comes from the need to be looking at a specific angle relative to the direction of the Sun, as illustrated in part (b). If two reflections of light occur within the water drop, another “secondary” rainbow is produced. This rare event produces an arc that lies above the primary rainbow arc, as in part (c), and produces colors in the reverse order of the primary rainbow, with red at the lowest angle and violet at the largest angle.



A ray of light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.



(a) (b) (c)

(a) Different colors emerge in different directions, and so you must look at different locations to see the various colors of a rainbow. (b) The arc of a rainbow results from the fact that a line between the observer and any point on the arc must make the correct angle with the parallel rays of sunlight for the observer to receive the refracted rays. (c) Double rainbow. (credit c: modification of work by “Nicholas”/Wikimedia Commons)

Dispersion may produce beautiful rainbows, but it can cause problems in optical systems. White light used to transmit messages in a fiber is dispersed, spreading out in time and eventually overlapping with other messages. Since a laser produces a nearly pure wavelength, its light experiences little dispersion, an advantage over white light for transmission of information. In contrast, dispersion of electromagnetic waves coming to us from outer space can be used to determine the amount of matter they pass through.

Summary

- The spreading of white light into its full spectrum of wavelengths is called dispersion.
- Rainbows are produced by a combination of refraction and reflection, and involve the dispersion of sunlight into a continuous distribution of colors.
- Dispersion produces beautiful rainbows but also causes problems in certain optical systems.

Conceptual Questions

Exercise:

Problem:

Is it possible that total internal reflection plays a role in rainbows? Explain in terms of indices of refraction and angles, perhaps referring to that shown below. Some of us have seen the formation of a double rainbow; is it physically possible to observe a triple rainbow?

(credit: "Chad"/Flickr)

Exercise:

Problem:

A high-quality diamond may be quite clear and colorless, transmitting all visible wavelengths with little absorption. Explain how it can sparkle with flashes of brilliant color when illuminated by white light.

Solution:

In addition to total internal reflection, rays that refract into and out of diamond crystals are subject to dispersion due to varying values of n across the spectrum, resulting in a sparkling display of colors.

Problems

Exercise:

Problem:

(a) What is the ratio of the speed of red light to violet light in diamond, based on [link]? (b) What is this ratio in polystyrene? (c) Which is more dispersive?

Exercise:

Problem:

A beam of white light goes from air into water at an incident angle of 75.0°. At what angles are the red (660 nm) and violet (410 nm) parts of the light refracted?

Solution:

46.5° for red, 46.0° for violet

Exercise:

Problem:

By how much do the critical angles for red (660 nm) and violet (410 nm) light differ in a diamond surrounded by air?

Exercise:

Problem:

(a) A narrow beam of light containing yellow (580 nm) and green (550 nm) wavelengths goes from polystyrene to air, striking the surface at a 30.0° incident angle. What is the angle between the colors when they emerge? (b) How far would they have to travel to be separated by 1.00 mm?

Solution:

a. 0.04°; b. 1.3 m

Exercise:

Problem:

A parallel beam of light containing orange (610 nm) and violet (410 nm) wavelengths goes from fused quartz to water, striking the surface between them at a 60.0° incident angle. What is the angle between the two colors in water?

Exercise:

Problem:

A ray of 610-nm light goes from air into fused quartz at an incident angle of 55.0°. At what incident angle must 470 nm light enter flint glass to have the same angle of refraction?

Solution:

72.8°

Exercise:

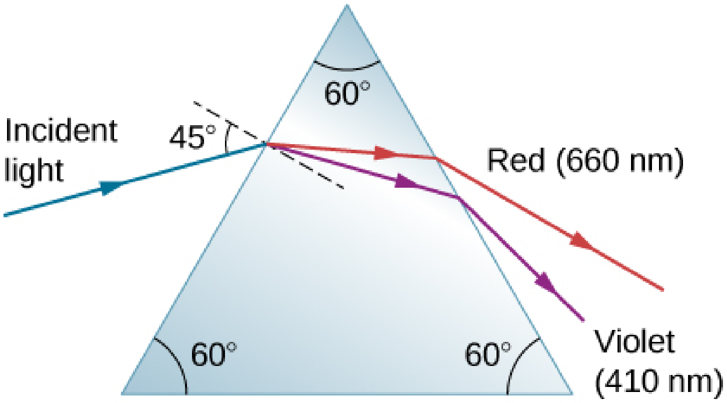
Problem:

A narrow beam of light containing red (660 nm) and blue (470 nm) wavelengths travels from air through a 1.00-cm-thick flat piece of crown glass and back to air again. The beam strikes at a 30.0° incident angle. (a) At what angles do the two colors emerge? (b) By what distance are the red and blue separated when they emerge?

Exercise:

Problem:

A narrow beam of white light enters a prism made of crown glass at a 45.0° incident angle, as shown below. At what angles, θ_R and θ_V, do the red (660 nm) and violet (410 nm) components of the light emerge from the prism?



Solution:

53.5° for red, 55.2° for violet

Glossary

dispersion
spreading of light into its spectrum of wavelengths

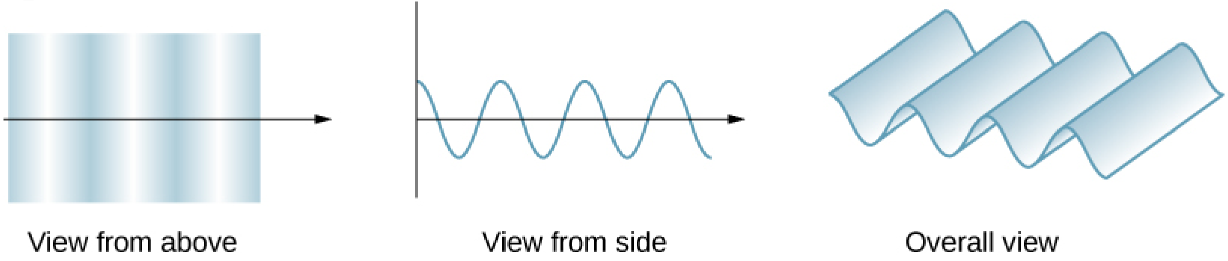
Huygens’s Principle

By the end of this section, you will be able to:

- Describe Huygens’s principle
- Use Huygens’s principle to explain the law of reflection
- Use Huygens’s principle to explain the law of refraction
- Use Huygens’s principle to explain diffraction

So far in this chapter, we have been discussing optical phenomena using the ray model of light. However, some phenomena require analysis and explanations based on the wave characteristics of light. This is particularly true when the wavelength is not negligible compared to the dimensions of an optical device, such as a slit in the case of *diffraction*. Huygens’s principle is an indispensable tool for this analysis.

[link] shows how a transverse wave looks as viewed from above and from the side. A light wave can be imagined to propagate like this, although we do not actually see it wiggling through space. From above, we view the wave fronts (or wave crests) as if we were looking down on ocean waves. The side view would be a graph of the electric or magnetic field. The view from above is perhaps more useful in developing concepts about **wave optics**.

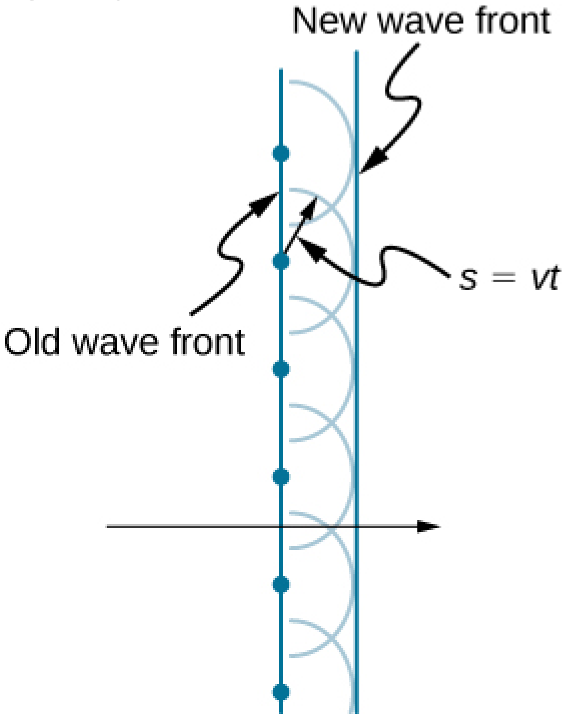


A transverse wave, such as an electromagnetic light wave, as viewed from above and from the side. The direction of propagation is perpendicular to the wave fronts (or wave crests) and is represented by a ray.

The Dutch scientist Christiaan Huygens (1629–1695) developed a useful technique for determining in detail how and where waves propagate.

Starting from some known position, **Huygens’s principle** states that every point on a wave front is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wave front is tangent to all of the wavelets.

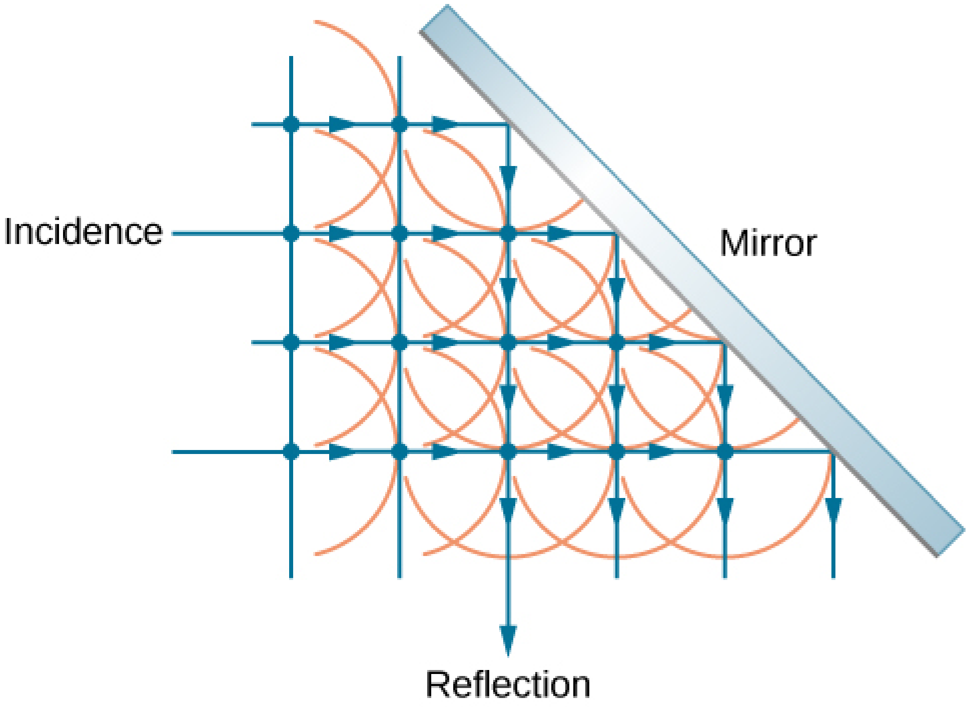
[link] shows how Huygens’s principle is applied. A wave front is the long edge that moves, for example, with the crest or the trough. Each point on the wave front emits a semicircular wave that moves at the propagation speed *v*. We can draw these wavelets at a time *t* later, so that they have moved a distance $s = vt$. The new wave front is a plane tangent to the wavelets and is where we would expect the wave to be a time *t* later. Huygens’s principle works for all types of waves, including water waves, sound waves, and light waves. It is useful not only in describing how light waves propagate but also in explaining the laws of reflection and refraction. In addition, we will see that Huygens’s principle tells us how and where light rays interfere.



Huygens’s principle applied to a straight wave front. Each point on the wave front emits a semicircular wavelet that moves a distance $s = vt$. The new wave front is a line tangent to the wavelets.

Reflection

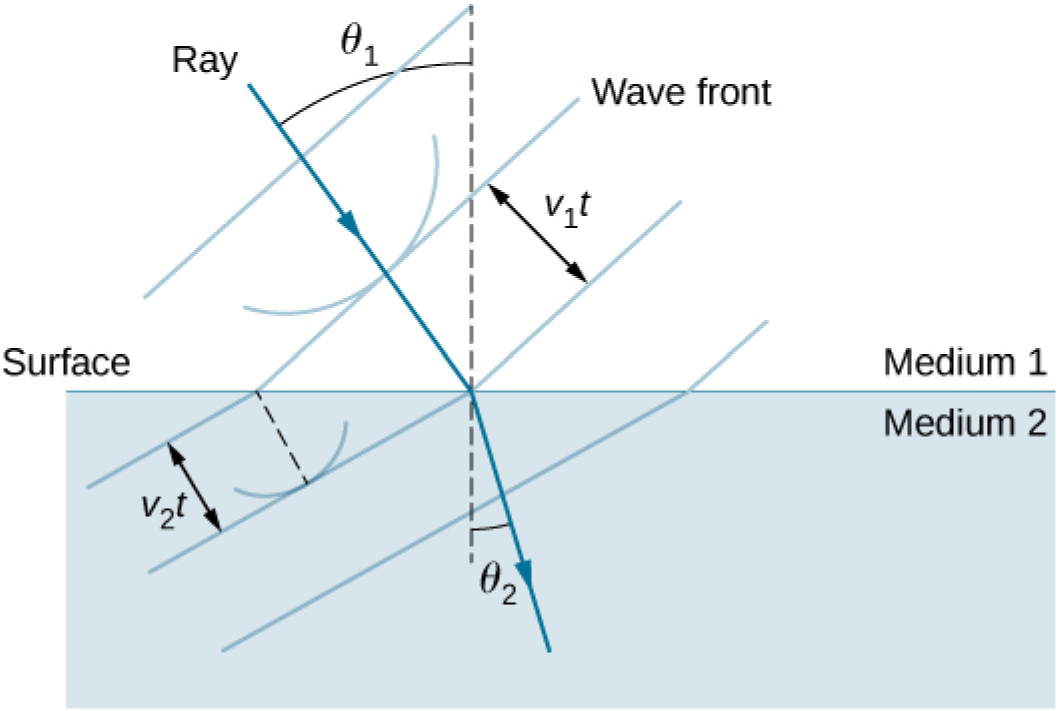
[link] shows how a mirror reflects an incoming wave at an angle equal to the incident angle, verifying the law of reflection. As the wave front strikes the mirror, wavelets are first emitted from the left part of the mirror and then from the right. The wavelets closer to the left have had time to travel farther, producing a wave front traveling in the direction shown.



Huygens’s principle applied to a plane wave front striking a mirror. The wavelets shown were emitted as each point on the wave front struck the mirror. The tangent to these wavelets shows that the new wave front has been reflected at an angle equal to the incident angle. The direction of propagation is perpendicular to the wave front, as shown by the downward-pointing arrows.

Refraction

The law of refraction can be explained by applying Huygens's principle to a wave front passing from one medium to another ([link]). Each wavelet in the figure was emitted when the wave front crossed the interface between the media. Since the speed of light is smaller in the second medium, the waves do not travel as far in a given time, and the new wave front changes direction as shown. This explains why a ray changes direction to become closer to the perpendicular when light slows down. Snell's law can be derived from the geometry in [link] ([link]).



Huygens's principle applied to a plane wave front traveling from one medium to another, where its speed is less. The ray bends toward the perpendicular, since the wavelets have a lower speed in the second medium.

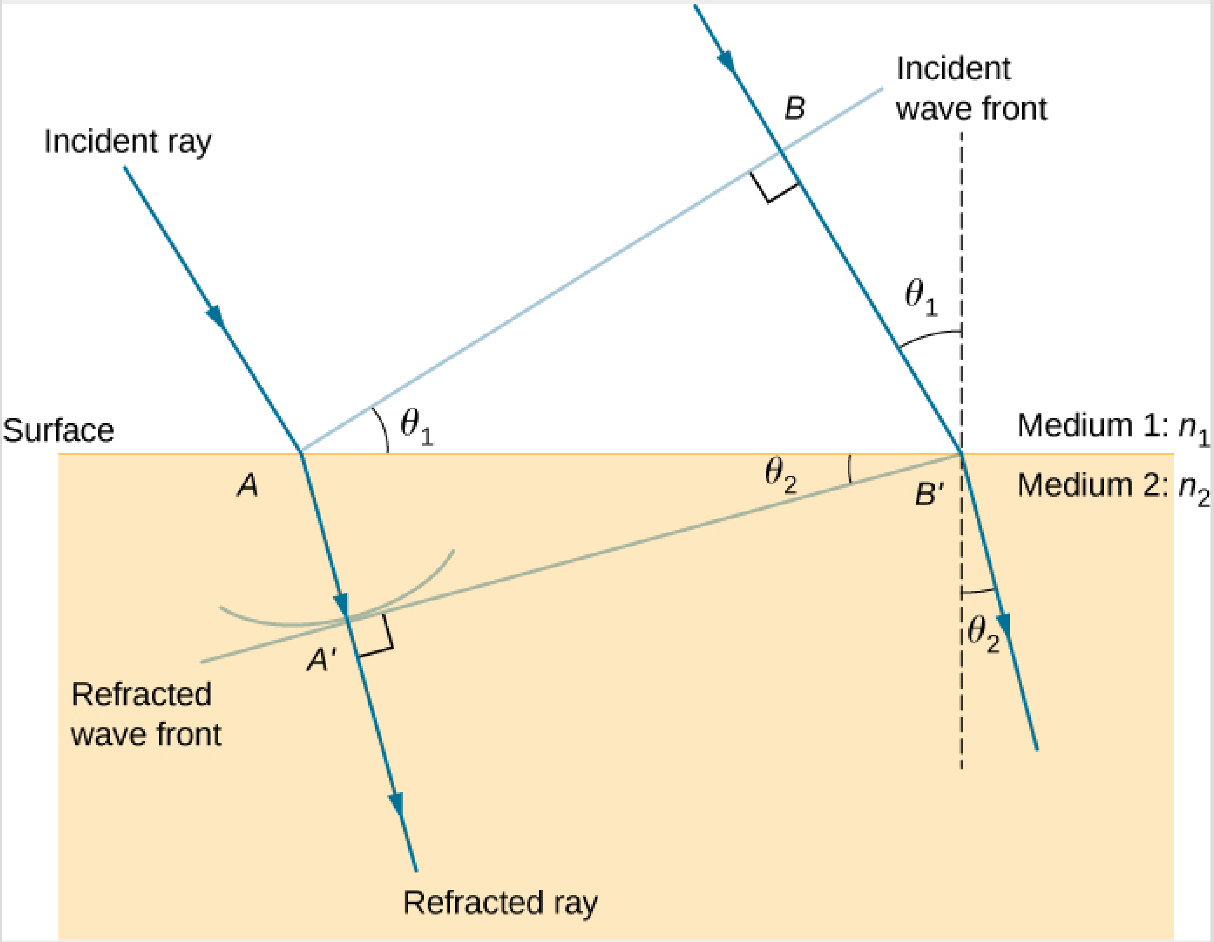
Example:

Deriving the Law of Refraction

By examining the geometry of the wave fronts, derive the law of refraction.

Strategy

Consider [link], which expands upon [link]. It shows the incident wave front just reaching the surface at point A, while point B is still well within medium 1. In the time Δt it takes for a wavelet from B to reach $B\prime$ on the surface at speed $v_1 = c/n_1$, a wavelet from A travels into medium 2 a distance of $AA\prime = v_2\Delta t$, where $v_2 = c/n_2$. Note that in this example, v_2 is slower than v_1 because $n_1 < n_2$.



Geometry of the law of refraction from medium 1 to medium 2.

Solution
The segment on the surface AB' is shared by both the triangle ABB' inside medium 1 and the triangle $AA'B'$ inside medium 2. Note that from the geometry, the angle $\angle BAB'$ is equal to the angle of incidence, θ_1. Similarly, $\angle AB'A'$ is θ_2.
The length of AB' is given in two ways as

Equation:

$$AB' = \frac{BB'}{\sin\theta_1} = \frac{AA'}{\sin\theta_2}.$$

Inverting the equation and substituting $AA' = c\Delta t/n_2$ from above and similarly $BB' = c\Delta t/n_1$, we obtain

Equation:

$$\frac{\sin\theta_1}{c\Delta t/n_1} = \frac{\sin\theta_2}{c\Delta t/n_2}.$$

Cancellation of $c\Delta t$ allows us to simplify this equation into the familiar form

Equation:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

Significance
Although the law of refraction was established experimentally by Snell and stated in Refraction, its derivation here requires Huygens's principle and the understanding that the speed of light is different in different media.

Note:
Exercise:

Problem:

Check Your Understanding In [link], we had $n_1 < n_2$. If n_2 were decreased such that $n_1 > n_2$ and the speed of light in medium 2 is faster than in medium 1, what would happen to the length of $AA\prime$? What would happen to the wave front $A\prime B\prime$ and the direction of the refracted ray?

Solution:

$AA\prime$ becomes longer, $A\prime B\prime$ tilts further away from the surface, and the refracted ray tilts away from the normal.

Note:
This applet by Walter Fendt shows an animation of reflection and refraction using Huygens’s wavelets while you control the parameters. Be sure to click on “Next step” to display the wavelets. You can see the reflected and refracted wave fronts forming.

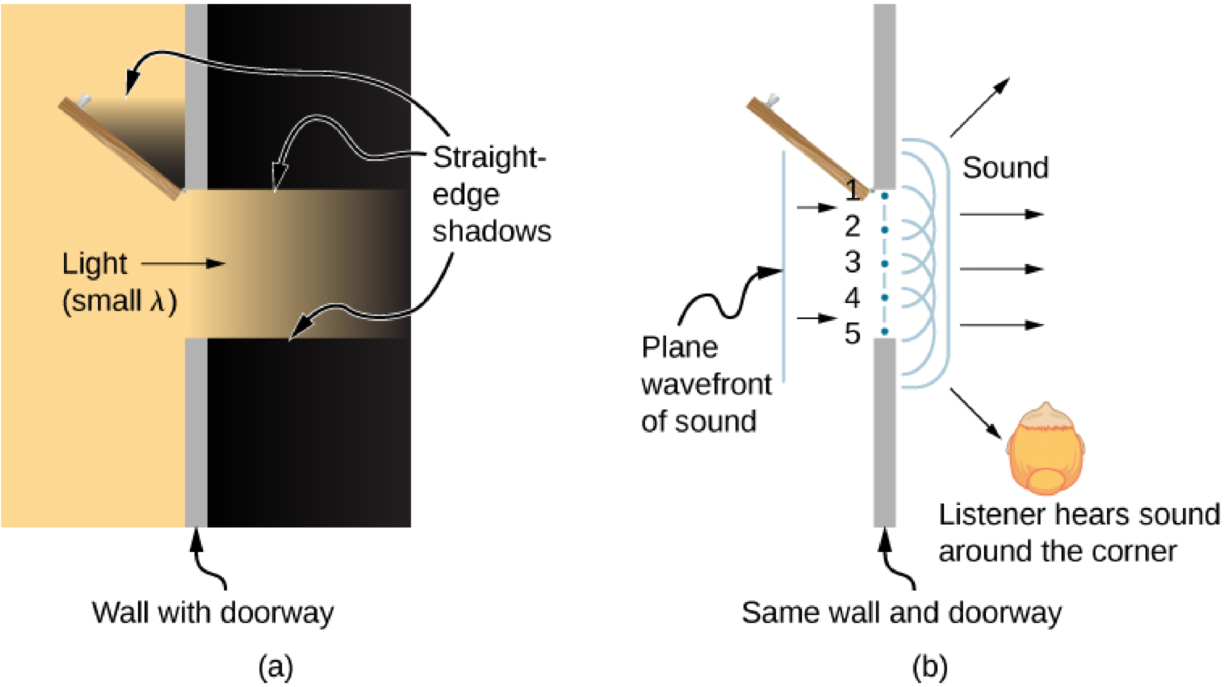
Diffraction

What happens when a wave passes through an opening, such as light shining through an open door into a dark room? For light, we observe a sharp shadow of the doorway on the floor of the room, and no visible light bends around corners into other parts of the room. When sound passes through a door, we hear it everywhere in the room and thus observe that sound spreads out when passing through such an opening ([link]). What is the difference between the behavior of sound waves and light waves in this case? The answer is that light has very short wavelengths and acts like a ray. Sound has wavelengths on the order of the size of the door and bends around corners (for frequency of 1000 Hz,

Equation:

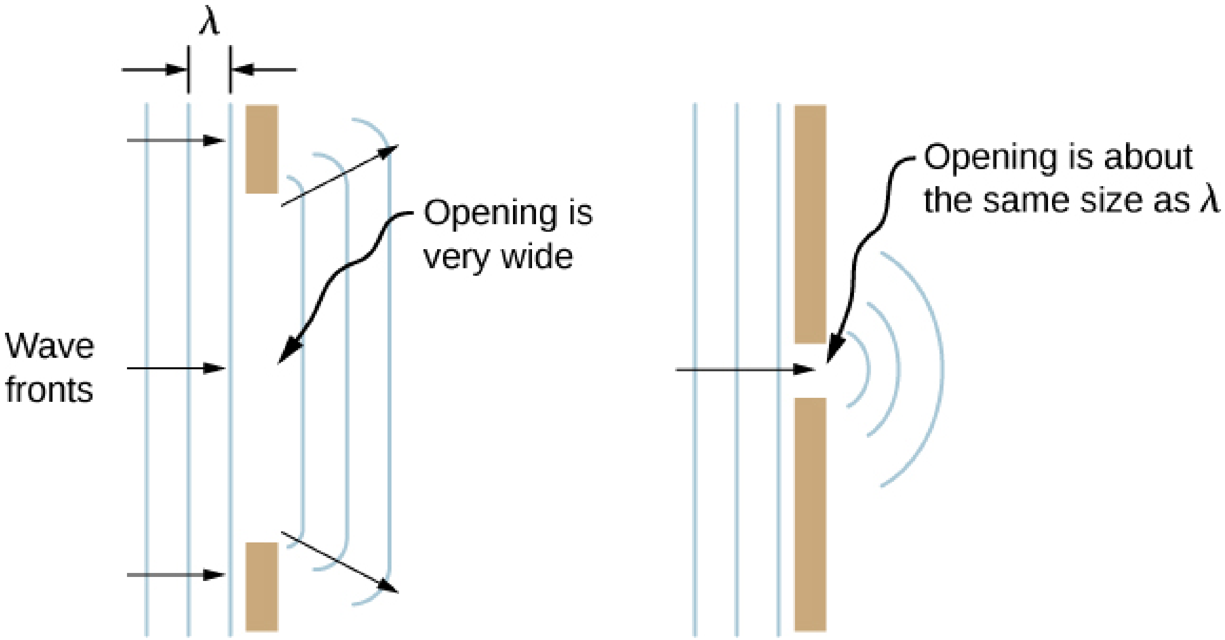
$$\lambda = \frac{c}{f} = \frac{330\,\text{m/s}}{1000\,\text{s}^{-1}} = 0.33\,\text{m},$$

about three times smaller than the width of the doorway).



(a) Light passing through a doorway makes a sharp outline on the floor. Since light’s wavelength is very small compared with the size of the door, it acts like a ray. (b) Sound waves bend into all parts of the room, a wave effect, because their wavelength is similar to the size of the door.

If we pass light through smaller openings such as slits, we can use Huygens’s principle to see that light bends as sound does ([link]). The bending of a wave around the edges of an opening or an obstacle is called diffraction. Diffraction is a wave characteristic and occurs for all types of waves. If diffraction is observed for some phenomenon, it is evidence that the phenomenon is a wave. Thus, the horizontal diffraction of the laser beam after it passes through the slits in [link] is evidence that light is a wave. You will learn about diffraction in much more detail in the chapter on Diffraction.



Huygens’s principle applied to a plane wave front striking an opening. The edges of the wave front bend after passing through the opening, a process called diffraction. The amount of bending is more extreme for a small opening, consistent with the fact that wave characteristics are most noticeable for interactions with objects about the same size as the wavelength.

Summary

- According to Huygens’s principle, every point on a wave front is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wave front is tangent to all of the wavelets.

- A mirror reflects an incoming wave at an angle equal to the incident angle, verifying the law of reflection.
- The law of refraction can be explained by applying Huygens’s principle to a wave front passing from one medium to another.
- The bending of a wave around the edges of an opening or an obstacle is called diffraction.

Conceptual Questions

Exercise:

Problem:

How do wave effects depend on the size of the object with which the wave interacts? For example, why does sound bend around the corner of a building while light does not?

Exercise:

Problem: Does Huygens’s principle apply to all types of waves?

Solution:

yes

Exercise:

Problem:

If diffraction is observed for some phenomenon, it is evidence that the phenomenon is a wave. Does the reverse hold true? That is, if diffraction is not observed, does that mean the phenomenon is not a wave?

Glossary

Huygens’s principle

every point on a wave front is a source of wavelets that spread out in the forward direction at the same speed as the wave itself; the new wave front is a plane tangent to all of the wavelets

wave optics

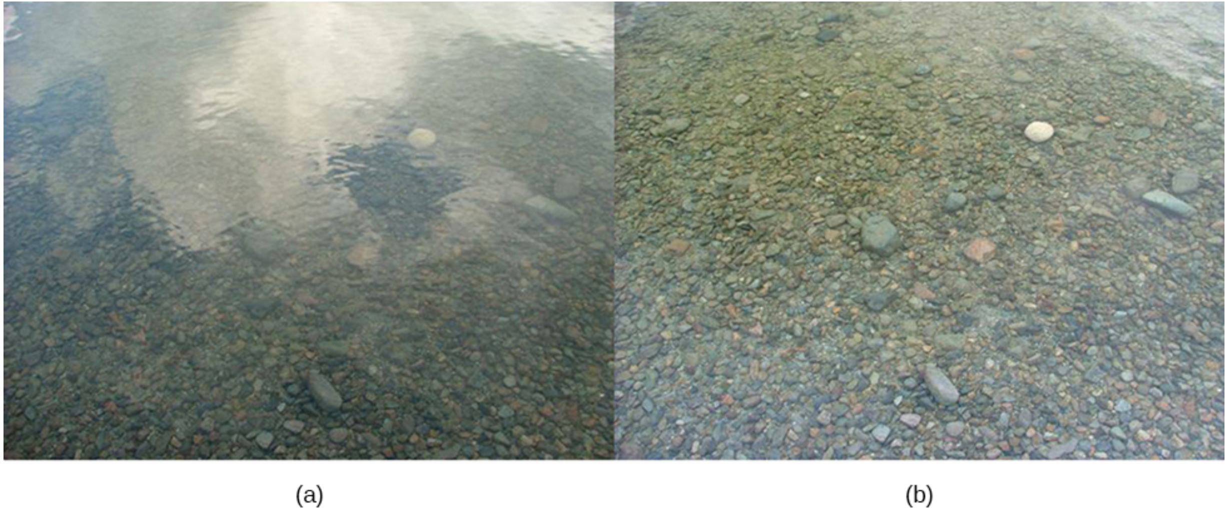
part of optics dealing with the wave aspect of light

Polarization

By the end of this section, you will be able to:

- Explain the change in intensity as polarized light passes through a polarizing filter
- Calculate the effect of polarization by reflection and Brewster's angle
- Describe the effect of polarization by scattering
- Explain the use of polarizing materials in devices such as LCDs

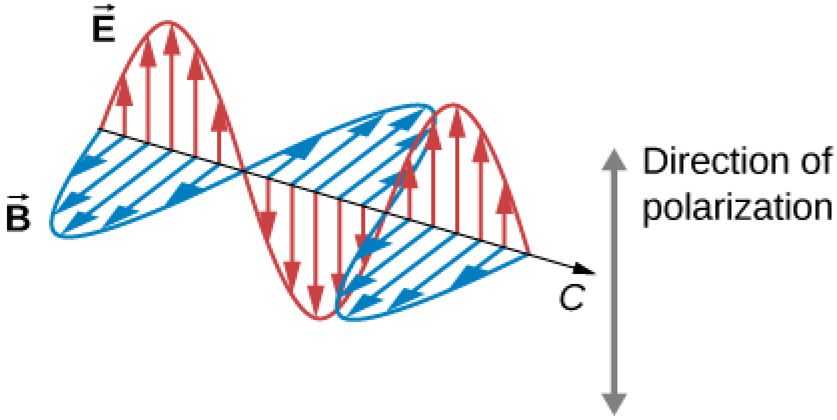
Polarizing sunglasses are familiar to most of us. They have a special ability to cut the glare of light reflected from water or glass ([link]). They have this ability because of a wave characteristic of light called polarization. What is polarization? How is it produced? What are some of its uses? The answers to these questions are related to the wave character of light.



These two photographs of a river show the effect of a polarizing filter in reducing glare in light reflected from the surface of water. Part (b) of this figure was taken with a polarizing filter and part (a) was not. As a result, the reflection of clouds and sky observed in part (a) is not observed in part (b). Polarizing sunglasses are particularly useful on snow and water. (credit a and credit b: modifications of work by "Amithshs"/Wikimedia Commons)

Malus's Law

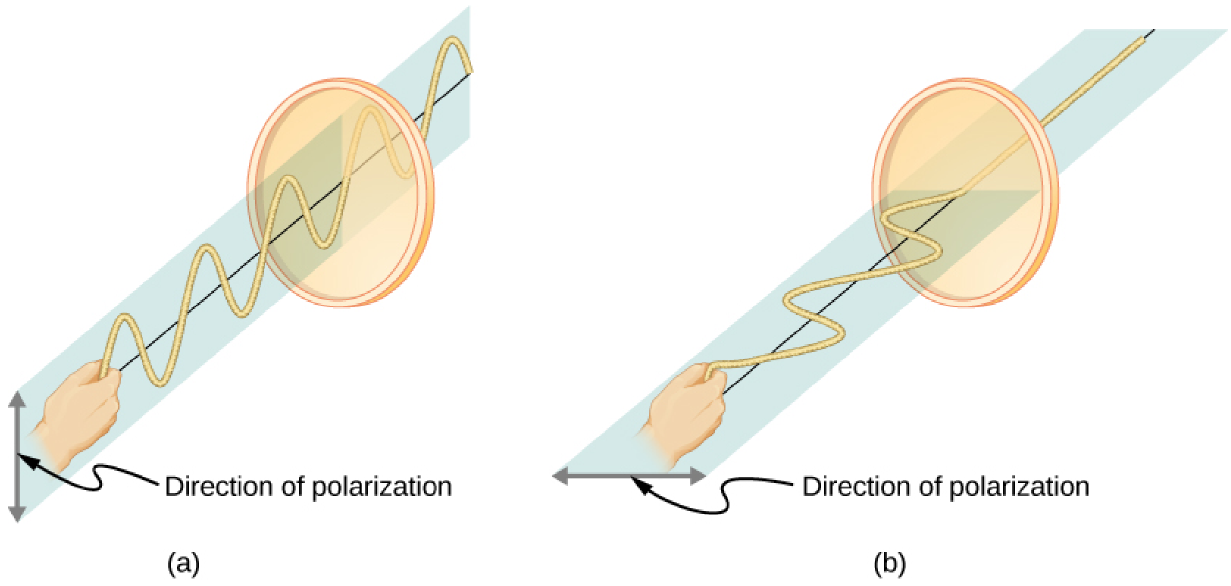
Light is one type of electromagnetic (EM) wave. As noted in the previous chapter on Electromagnetic Waves, EM waves are *transverse waves* consisting of varying electric and magnetic fields that oscillate perpendicular to the direction of propagation ([link]). However, in general, there are no specific directions for the oscillations of the electric and magnetic fields; they vibrate in any randomly oriented plane perpendicular to the direction of propagation. **Polarization** is the attribute that a wave's oscillations do have a definite direction relative to the direction of propagation of the wave. (This is not the same type of polarization as that discussed for the separation of charges.) Waves having such a direction are said to be **polarized**. For an EM wave, we define the **direction of polarization** to be the direction parallel to the electric field. Thus, we can think of the electric field arrows as showing the direction of polarization, as in [link].



An EM wave, such as light, is a transverse wave. The electric $(\vec{\mathbf{E}})$ and magnetic $(\vec{\mathbf{B}})$ fields are perpendicular to the direction of propagation. The direction of polarization of the wave is the direction of the electric field.

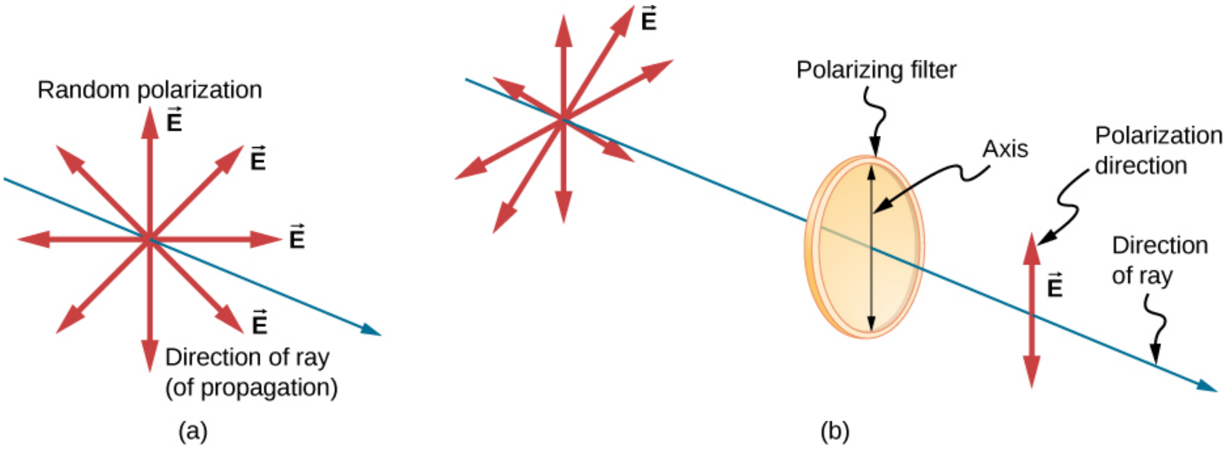
To examine this further, consider the transverse waves in the ropes shown in [link]. The oscillations in one rope are in a vertical plane and are said to be

vertically polarized. Those in the other rope are in a horizontal plane and are **horizontally polarized**. If a vertical slit is placed on the first rope, the waves pass through. However, a vertical slit blocks the horizontally polarized waves. For EM waves, the direction of the electric field is analogous to the disturbances on the ropes.



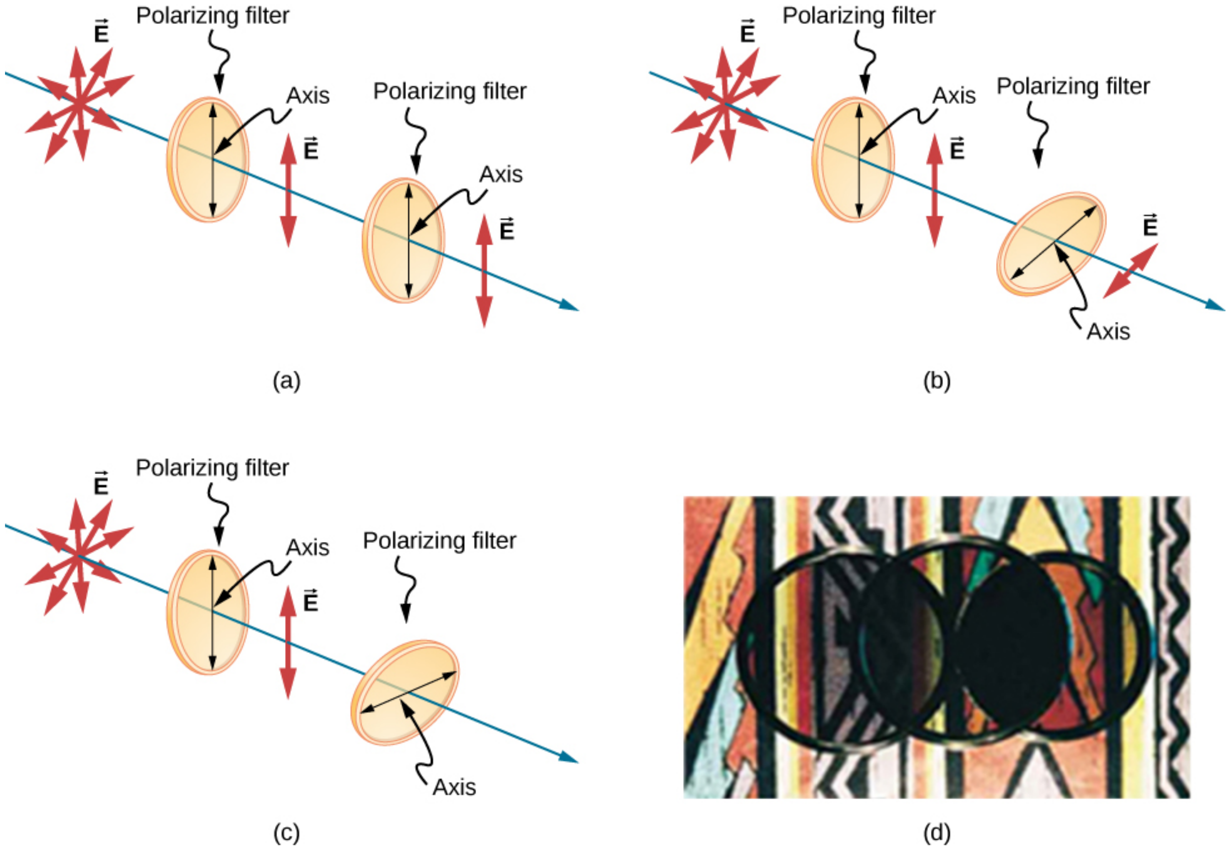
The transverse oscillations in one rope (a) are in a vertical plane, and those in the other rope (b) are in a horizontal plane. The first is said to be vertically polarized, and the other is said to be horizontally polarized. Vertical slits pass vertically polarized waves and block horizontally polarized waves.

The Sun and many other light sources produce waves that have the electric fields in random directions ([link](a)). Such light is said to be **unpolarized**, because it is composed of many waves with all possible directions of polarization. Polaroid materials—which were invented by the founder of the Polaroid Corporation, Edwin Land—act as a polarizing slit for light, allowing only polarization in one direction to pass through. Polarizing filters are composed of long molecules aligned in one direction. If we think of the molecules as many slits, analogous to those for the oscillating ropes, we can understand why only light with a specific polarization can get through. The axis of a polarizing filter is the direction along which the filter passes the electric field of an EM wave.



The slender arrow represents a ray of unpolarized light. The bold arrows represent the direction of polarization of the individual waves composing the ray. (a) If the light is unpolarized, the arrows point in all directions. (b) A polarizing filter has a polarization axis that acts as a slit passing through electric fields parallel to its direction. The direction of polarization of an EM wave is defined to be the direction of its electric field.

[link] shows the effect of two polarizing filters on originally unpolarized light. The first filter polarizes the light along its axis. When the axes of the first and second filters are aligned (parallel), then all of the polarized light passed by the first filter is also passed by the second filter. If the second polarizing filter is rotated, only the component of the light parallel to the second filter’s axis is passed. When the axes are perpendicular, no light is passed by the second filter.



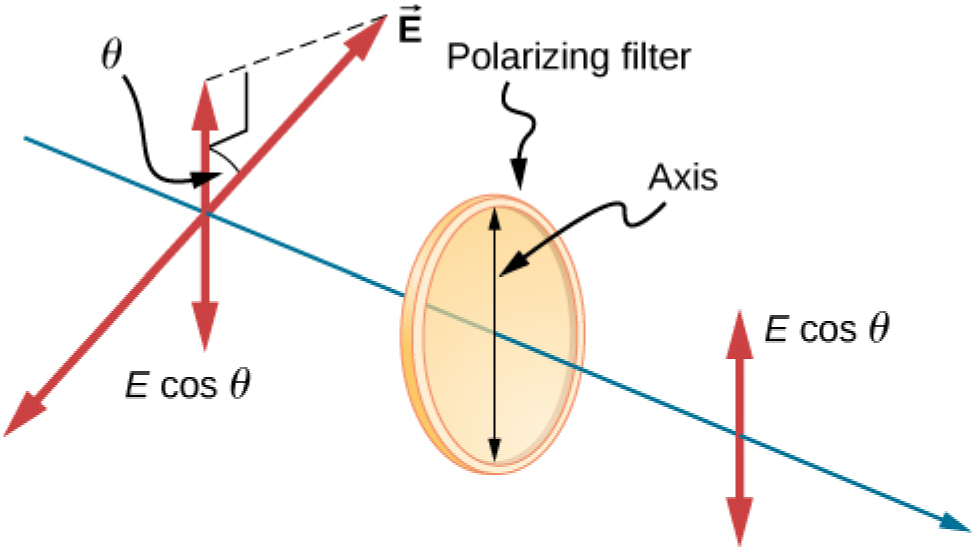
The effect of rotating two polarizing filters, where the first polarizes the light. (a) All of the polarized light is passed by the second polarizing filter, because its axis is parallel to the first. (b) As the second filter is rotated, only part of the light is passed. (c) When the second filter is perpendicular to the first, no light is passed. (d) In this photograph, a polarizing filter is placed above two others. Its axis is perpendicular to the filter on the right (dark area) and parallel to the filter on the left (lighter area). (credit d: modification of work by P.P. Urone)

Only the component of the EM wave parallel to the axis of a filter is passed. Let us call the angle between the direction of polarization and the axis of a filter θ. If the electric field has an amplitude E, then the transmitted part of the wave has an amplitude $E \cos \theta$ ([link]). Since the intensity of a wave is proportional to its amplitude squared, the intensity I of the transmitted wave is related to the incident wave by

Note:
Equation:

$$I = I_0 \cos^2 \theta$$

where I_0 is the intensity of the polarized wave before passing through the filter. This equation is known as **Malus's law**.



A polarizing filter transmits only the component of the wave parallel to its axis, reducing the intensity of any light not polarized parallel to its axis.

Note:
This Open Source Physics animation helps you visualize the electric field vectors as light encounters a polarizing filter. You can rotate the filter—

note that the angle displayed is in radians. You can also rotate the animation for 3D visualization.

Example:

Calculating Intensity Reduction by a Polarizing Filter

What angle is needed between the direction of polarized light and the axis of a polarizing filter to reduce its intensity by 90.0%?

Strategy

When the intensity is reduced by 90.0%, it is 10.0% or 0.100 times its original value. That is, $I = 0.100\, I_0$. Using this information, the equation $I = I_0 \cos^2 \theta$ can be used to solve for the needed angle.

Solution

Solving the equation $I = I_0 \cos^2 \theta$ for $\cos \theta$ and substituting with the relationship between I and I_0 gives

Equation:

$$\cos \theta = \sqrt{\frac{I}{I_0}} = \sqrt{\frac{0.100\, I_0}{I_0}} = 0.3162.$$

Solving for θ yields

Equation:

$$\theta = \cos^{-1} 0.3162 = 71.6^\circ.$$

Significance

A fairly large angle between the direction of polarization and the filter axis is needed to reduce the intensity to 10.0% of its original value. This seems reasonable based on experimenting with polarizing films. It is interesting that at an angle of 45°, the intensity is reduced to 50% of its original value. Note that 71.6° is 18.4° from reducing the intensity to zero, and that at an angle of 18.4°, the intensity is reduced to 90.0% of its original value, giving evidence of symmetry.

Note:
Exercise:

Problem:

Check Your Understanding Although we did not specify the direction in [link], let’s say the polarizing filter was rotated clockwise by 71.6° to reduce the light intensity by 90.0%. What would be the intensity reduction if the polarizing filter were rotated counterclockwise by 71.6°?

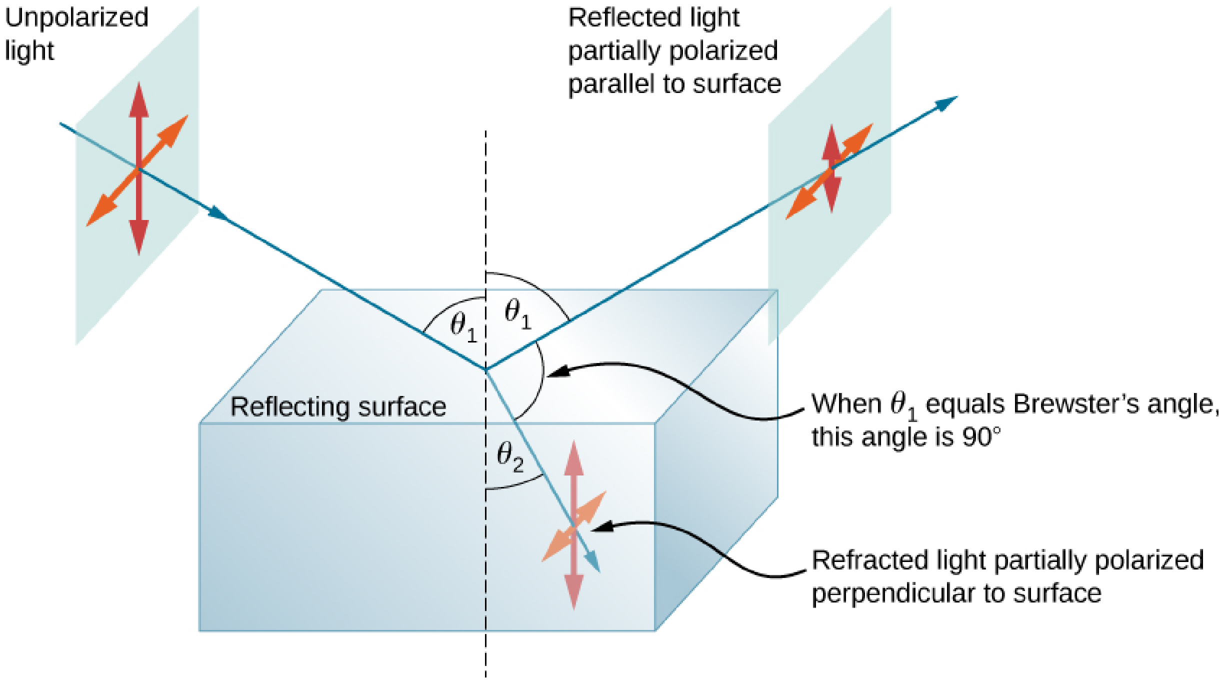
Solution:

also 90.0%

Polarization by Reflection

By now, you can probably guess that polarizing sunglasses cut the glare in reflected light, because that light is polarized. You can check this for yourself by holding polarizing sunglasses in front of you and rotating them while looking at light reflected from water or glass. As you rotate the sunglasses, you will notice the light gets bright and dim, but not completely black. This implies the reflected light is partially polarized and cannot be completely blocked by a polarizing filter.

[link] illustrates what happens when unpolarized light is reflected from a surface. Vertically polarized light is preferentially refracted at the surface, so the reflected light is left more horizontally polarized. The reasons for this phenomenon are beyond the scope of this text, but a convenient mnemonic for remembering this is to imagine the polarization direction to be like an arrow. Vertical polarization is like an arrow perpendicular to the surface and is more likely to stick and not be reflected. Horizontal polarization is like an arrow bouncing on its side and is more likely to be reflected. Sunglasses with vertical axes thus block more reflected light than unpolarized light from other sources.


Unpolarized light
Reflected light partially polarized parallel to surface
θ_1
θ_1
Reflecting surface
θ_2
When θ_1 equals Brewster's angle, this angle is 90°
Refracted light partially polarized perpendicular to surface

Polarization by reflection. Unpolarized light has equal amounts of vertical and horizontal polarization. After interaction with a surface, the vertical components are preferentially absorbed or refracted, leaving the reflected light more horizontally polarized. This is akin to arrows striking on their sides and bouncing off, whereas arrows striking on their tips go into the surface.

Since the part of the light that is not reflected is refracted, the amount of polarization depends on the indices of refraction of the media involved. It can be shown that reflected light is completely polarized at an angle of reflection θ_{b} given by

Note:
Equation:

$$\tan\theta_{\text{b}} = \frac{n_2}{n_1}$$

where n_1 is the medium in which the incident and reflected light travel and n_2 is the index of refraction of the medium that forms the interface that reflects the light. This equation is known as **Brewster's law** and θ_b is known as **Brewster's angle**, named after the nineteenth-century Scottish physicist who discovered them.

Note:
This Open Source Physics animation shows incident, reflected, and refracted light as rays and EM waves. Try rotating the animation for 3D visualization and also change the angle of incidence. Near Brewster's angle, the reflected light becomes highly polarized.

Example:

Calculating Polarization by Reflection

(a) At what angle will light traveling in air be completely polarized horizontally when reflected from water? (b) From glass?

Strategy

All we need to solve these problems are the indices of refraction. Air has $n_1 = 1.00$, water has $n_2 = 1.333$, and crown glass has $n'_2 = 1.520$. The equation $\tan\theta_\text{b} = \frac{n_2}{n_1}$ can be directly applied to find θ_b in each case.

Solution

a. Putting the known quantities into the equation
   Equation:

$$\tan\theta_\text{b} = \frac{n_2}{n_1}$$

   gives
   Equation:

$$\tan\theta_\text{b} = \frac{n_2}{n_1} = \frac{1.333}{1.00} = 1.333.$$

Solving for the angle θ_{b} yields
Equation:

$$\theta_{\mathrm{b}} = \tan^{-1} 1.333 = 53.1^{\circ}.$$

b. Similarly, for crown glass and air,
Equation:

$$\tan\theta'_{\mathrm{b}} = \frac{n'_2}{n_1} = \frac{1.520}{1.00} = 1.52.$$

Thus,
Equation:

$$\theta'_{\mathrm{b}} = \tan^{-1} 1.52 = 56.7^{\circ}.$$

Significance
Light reflected at these angles could be completely blocked by a good polarizing filter held with its axis vertical. Brewster’s angle for water and air are similar to those for glass and air, so that sunglasses are equally effective for light reflected from either water or glass under similar circumstances. Light that is not reflected is refracted into these media. Therefore, at an incident angle equal to Brewster’s angle, the refracted light is slightly polarized vertically. It is not completely polarized vertically, because only a small fraction of the incident light is reflected, so a significant amount of horizontally polarized light is refracted.

Note:
Exercise:

Problem:

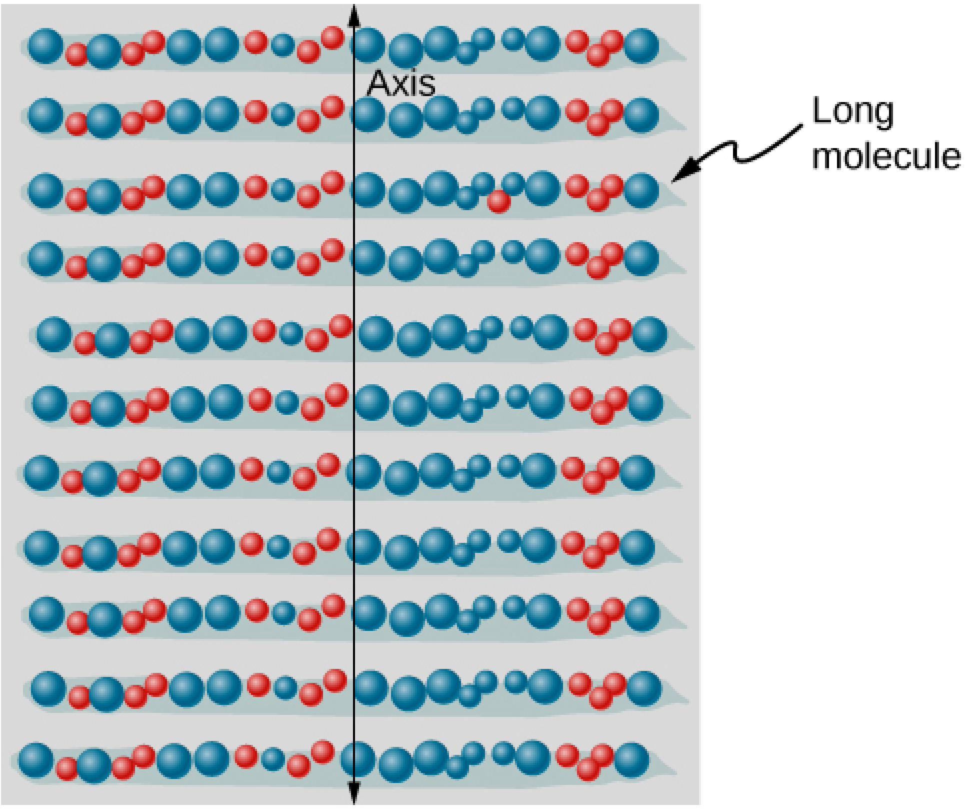
Check Your Understanding What happens at Brewster’s angle if the original incident light is already 100% vertically polarized?

Solution:

There will be only refraction but no reflection.

Atomic Explanation of Polarizing Filters

Polarizing filters have a polarization axis that acts as a slit. This slit passes EM waves (often visible light) that have an electric field parallel to the axis. This is accomplished with long molecules aligned perpendicular to the axis, as shown in [link].



Long molecules are aligned perpendicular to the axis of a polarizing filter. In an EM wave, the component of the electric field perpendicular to these molecules passes through the filter, whereas the component parallel to the molecules is absorbed.

[link] illustrates how the component of the electric field parallel to the long molecules is absorbed. An EM wave is composed of oscillating electric and magnetic fields. The electric field is strong compared with the magnetic field and is more effective in exerting force on charges in the molecules. The most affected charged particles are the electrons, since electron masses are small. If an electron is forced to oscillate, it can absorb energy from the EM wave. This reduces the field in the wave and, hence, reduces its intensity. In long molecules, electrons can more easily oscillate parallel to the molecule than in the perpendicular direction. The electrons are bound to the molecule and are more restricted in their movement perpendicular to the molecule. Thus, the electrons can absorb EM waves that have a component

of their electric field parallel to the molecule. The electrons are much less responsive to electric fields perpendicular to the molecule and allow these fields to pass. Thus, the axis of the polarizing filter is perpendicular to the length of the molecule.

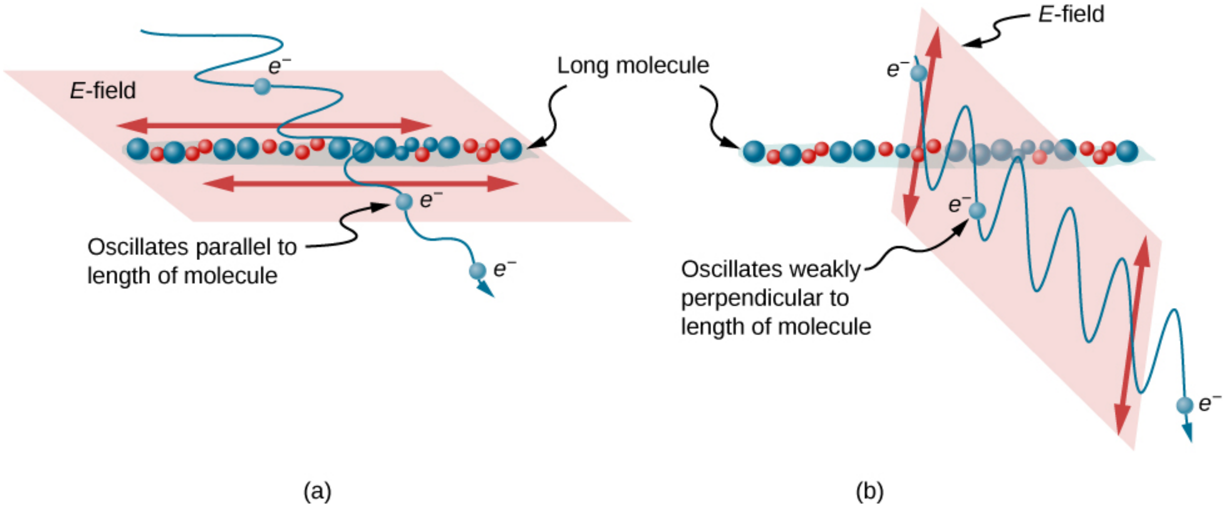
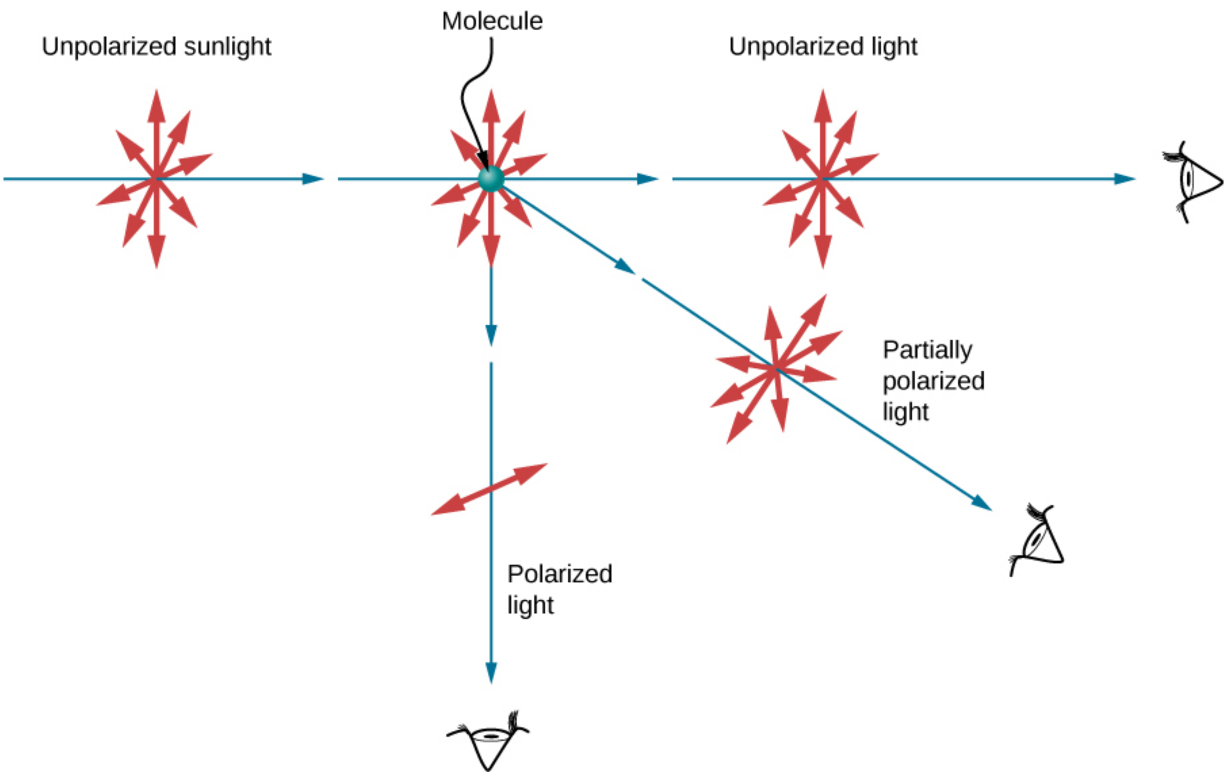

E-field
e⁻
Long molecule
E-field
e⁻
e⁻
Oscillates parallel to length of molecule
e⁻
e⁻
Oscillates weakly perpendicular to length of molecule
e⁻
(a)
(b)

Diagram of an electron in a long molecule oscillating parallel to the molecule. The oscillation of the electron absorbs energy and reduces the intensity of the component of the EM wave that is parallel to the molecule.

Polarization by Scattering

If you hold your polarizing sunglasses in front of you and rotate them while looking at blue sky, you will see the sky get bright and dim. This is a clear indication that light scattered by air is partially polarized. [link] helps illustrate how this happens. Since light is a transverse EM wave, it vibrates the electrons of air molecules perpendicular to the direction that it is traveling. The electrons then radiate like small antennae. Since they are oscillating perpendicular to the direction of the light ray, they produce EM radiation that is polarized perpendicular to the direction of the ray. When viewing the light along a line perpendicular to the original ray, as in the figure, there can be no polarization in the scattered light parallel to the original ray, because that would require the original ray to be a longitudinal wave. Along other directions, a component of the other polarization can be projected along the line of sight, and the scattered light is only partially polarized. Furthermore, multiple scattering can bring light to your eyes from other directions and can contain different polarizations.



Polarization by scattering. Unpolarized light scattering from air molecules shakes their electrons perpendicular to the direction of the original ray. The scattered light therefore has a polarization perpendicular to the original direction and none parallel to the original direction.

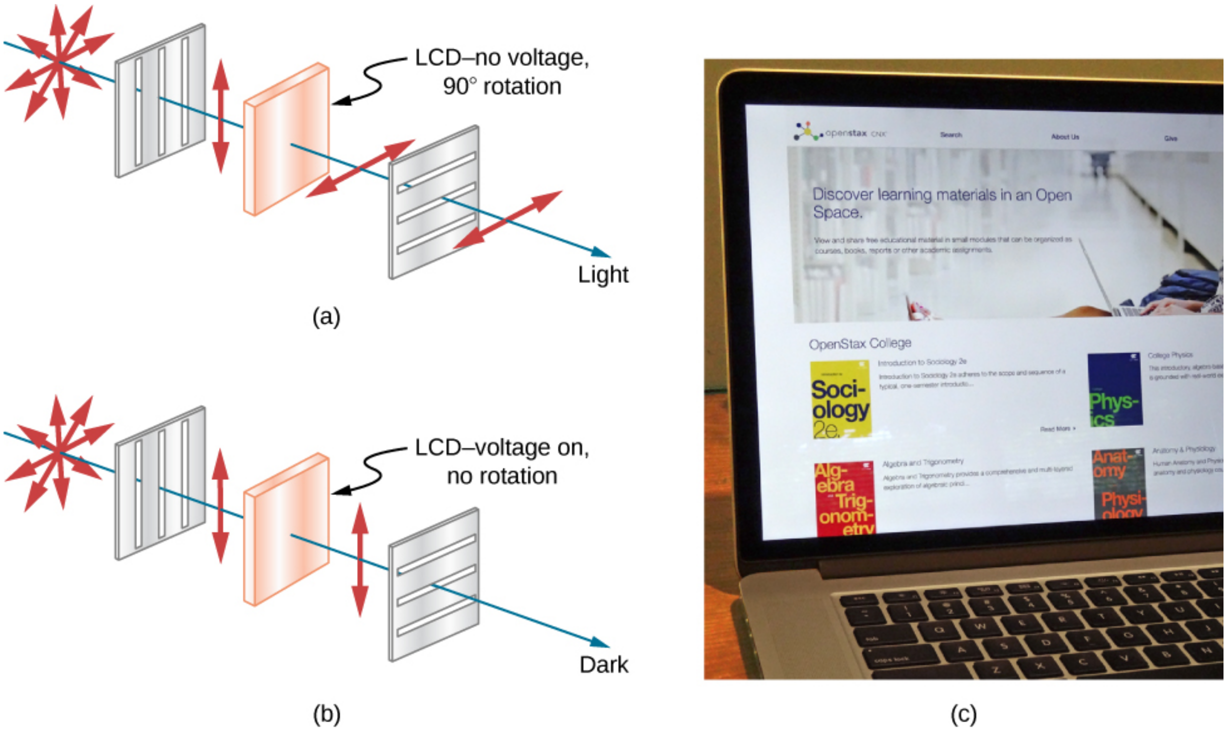
Photographs of the sky can be darkened by polarizing filters, a trick used by many photographers to make clouds brighter by contrast. Scattering from other particles, such as smoke or dust, can also polarize light. Detecting polarization in scattered EM waves can be a useful analytical tool in determining the scattering source.

A range of optical effects are used in sunglasses. Besides being polarizing, sunglasses may have colored pigments embedded in them, whereas others use either a nonreflective or reflective coating. A recent development is photochromic lenses, which darken in the sunlight and become clear indoors. Photochromic lenses are embedded with organic microcrystalline molecules that change their properties when exposed to UV in sunlight, but become clear in artificial lighting with no UV.

Liquid Crystals and Other Polarization Effects in Materials

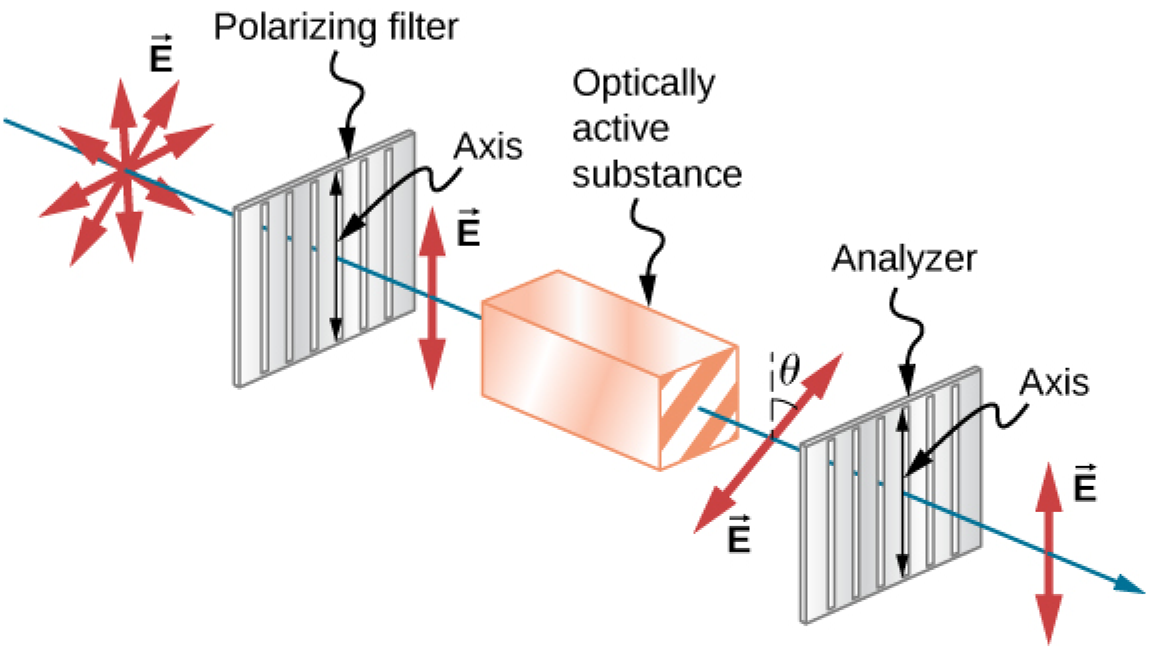
Although you are undoubtedly aware of liquid crystal displays (LCDs) found in watches, calculators, computer screens, cellphones, flat screen televisions, and many other places, you may not be aware that they are based on polarization. Liquid crystals are so named because their molecules can be aligned even though they are in a liquid. Liquid crystals have the property that they can rotate the polarization of light passing through them by 90°. Furthermore, this property can be turned off by the application of a voltage, as illustrated in [link]. It is possible to manipulate this characteristic quickly and in small, well-defined regions to create the contrast patterns we see in so many LCD devices.

In flat screen LCD televisions, a large light is generated at the back of the TV. The light travels to the front screen through millions of tiny units called pixels (picture elements). One of these is shown in [link](a) and (b). Each unit has three cells, with red, blue, or green filters, each controlled independently. When the voltage across a liquid crystal is switched off, the liquid crystal passes the light through the particular filter. We can vary the picture contrast by varying the strength of the voltage applied to the liquid crystal.


LCD–no voltage, 90° rotation
Light
(a)
LCD–voltage on, no rotation
Dark
(b)
Discover learning materials in an Open Space.
OpenStax College
Sociology 2e
Physics
Algebra and Trigonometry
Anatomy Physiology
(c)

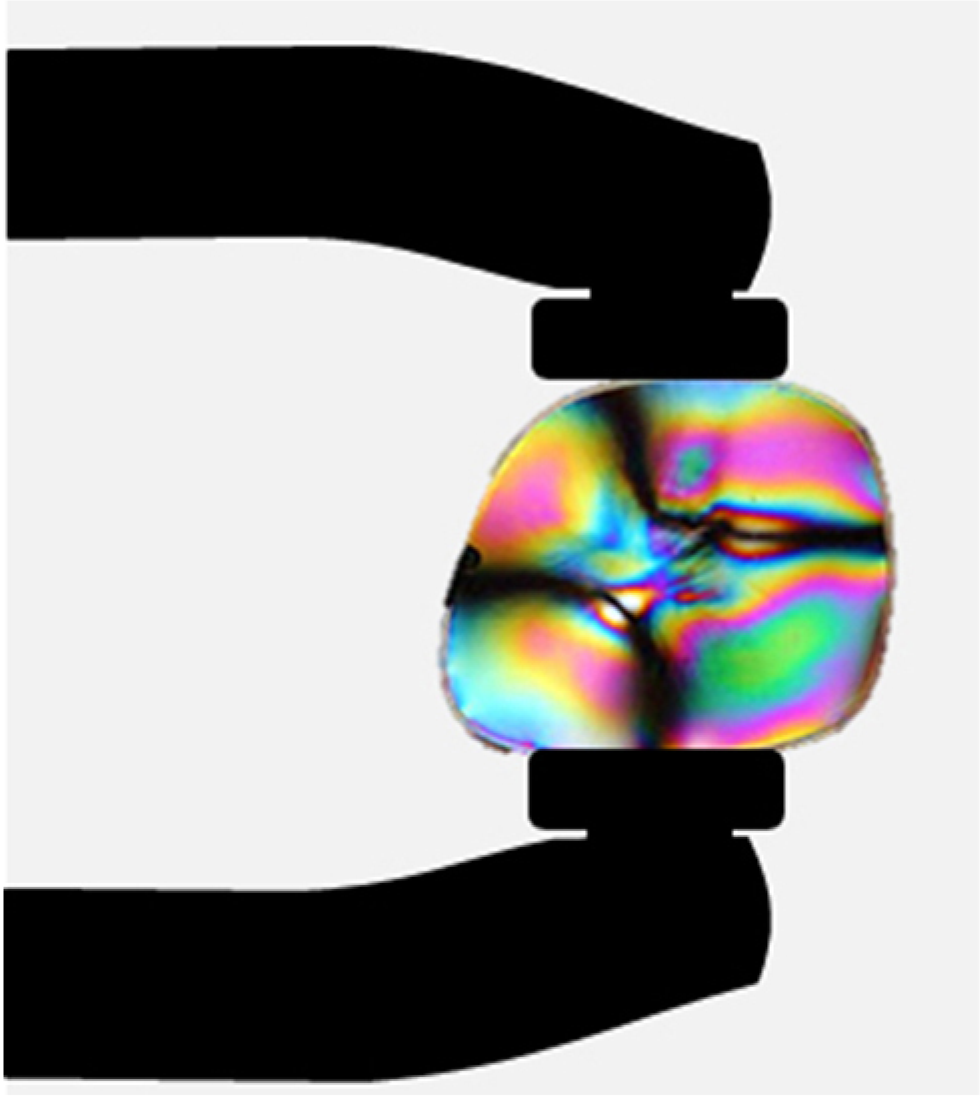
(a) Polarized light is rotated 90° by a liquid crystal and then passed by a polarizing filter that has its axis perpendicular to the direction of the original polarization. (b) When a voltage is applied to the liquid crystal, the polarized light is not rotated and is blocked by the filter, making the region dark in comparison with its surroundings. (c) LCDs can be made color specific, small, and fast enough to use in laptop computers and TVs. (credit c: modification of work by Jane Whitney)

Many crystals and solutions rotate the plane of polarization of light passing through them. Such substances are said to be **optically active**. Examples include sugar water, insulin, and collagen ([link]). In addition to depending on the type of substance, the amount and direction of rotation depend on several other factors. Among these is the concentration of the substance, the distance the light travels through it, and the wavelength of light. Optical activity is due to the asymmetrical shape of molecules in the substance, such as being helical. Measurements of the rotation of polarized light passing through substances can thus be used to measure concentrations, a standard technique for sugars. It can also give information on the shapes of molecules, such as proteins, and factors that affect their shapes, such as temperature and pH.



Optical activity is the ability of some substances to rotate the plane of polarization of light passing through them. The rotation is detected with a polarizing filter or analyzer.

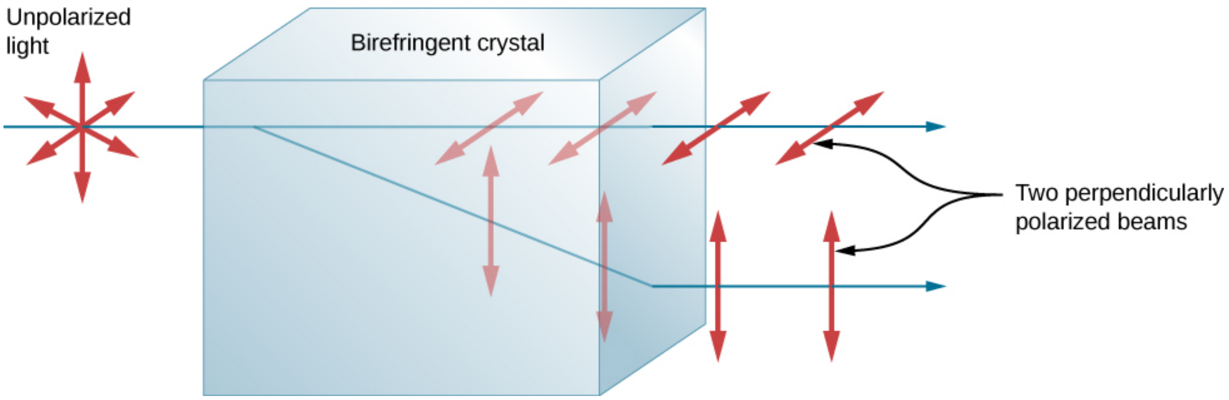
Glass and plastic become optically active when stressed: the greater the stress, the greater the effect. Optical stress analysis on complicated shapes can be performed by making plastic models of them and observing them through crossed filters, as seen in [link]. It is apparent that the effect depends on wavelength as well as stress. The wavelength dependence is sometimes also used for artistic purposes.



Optical stress analysis of a plastic lens placed between crossed polarizers. (credit: “Infopro”/Wikimedia Commons)

Another interesting phenomenon associated with polarized light is the ability of some crystals to split an unpolarized beam of light into two polarized beams. This occurs because the crystal has one value for the index of refraction of polarized light but a different value for the index of refraction of light polarized in the perpendicular direction, so that each component has its own angle of refraction. Such crystals are said to be

birefringent, and, when aligned properly, two perpendicularly polarized beams will emerge from the crystal ([link]). Birefringent crystals can be used to produce polarized beams from unpolarized light. Some birefringent materials preferentially absorb one of the polarizations. These materials are called dichroic and can produce polarization by this preferential absorption. This is fundamentally how polarizing filters and other polarizers work.



Birefringent materials, such as the common mineral calcite, split unpolarized beams of light into two with two different values of index of refraction.

Summary

- Polarization is the attribute that wave oscillations have a definite direction relative to the direction of propagation of the wave. The direction of polarization is defined to be the direction parallel to the electric field of the EM wave.
- Unpolarized light is composed of many rays having random polarization directions.
- Unpolarized light can be polarized by passing it through a polarizing filter or other polarizing material. The process of polarizing light decreases its intensity by a factor of 2.

- The intensity, I, of polarized light after passing through a polarizing filter is $I = I_0 \cos^2 \theta$, where I_0 is the incident intensity and θ is the angle between the direction of polarization and the axis of the filter.
- Polarization is also produced by reflection.
- Brewster's law states that reflected light is completely polarized at the angle of reflection θ_b, known as Brewster's angle.
- Polarization can also be produced by scattering.
- Several types of optically active substances rotate the direction of polarization of light passing through them.

Key Equations

Speed of light	$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$
Index of refraction	$n = \frac{c}{v}$
Law of reflection	$\theta_\text{r} = \theta_\text{i}$
Law of refraction (Snell's law)	$n_1 \sin \theta_1 = n_2 \sin \theta_2$
Critical angle	$\theta_\text{c} = \sin^{-1} \left(\frac{n_2}{n_1} \right) \text{ for } n_1 > n_2$
Malus's law	$I = I_0 \cos^2 \theta$
Brewster's law	$\tan \theta_\text{b} = \frac{n_2}{n_1}$

Conceptual Questions

Exercise:

Problem: Can a sound wave in air be polarized? Explain.

Solution:

No. Sound waves are not transverse waves.

Exercise:

Problem:

No light passes through two perfect polarizing filters with perpendicular axes. However, if a third polarizing filter is placed between the original two, some light can pass. Why is this? Under what circumstances does most of the light pass?

Exercise:

Problem:

Explain what happens to the energy carried by light that it is dimmed by passing it through two crossed polarizing filters.

Solution:

Energy is absorbed into the filters.

Exercise:

Problem:

When particles scattering light are much smaller than its wavelength, the amount of scattering is proportional to $\frac{1}{\lambda}$. Does this mean there is more scattering for small λ than large λ? How does this relate to the fact that the sky is blue?

Exercise:

Problem:

Using the information given in the preceding question, explain why sunsets are red.

Solution:

Sunsets are viewed with light traveling straight from the Sun toward us. When blue light is scattered out of this path, the remaining red light dominates the overall appearance of the setting Sun.

Exercise:

Problem:

When light is reflected at Brewster's angle from a smooth surface, it is 100% polarized parallel to the surface. Part of the light will be refracted into the surface. Describe how you would do an experiment to determine the polarization of the refracted light. What direction would you expect the polarization to have and would you expect it to be 100%?

Exercise:

Problem:

If you lie on a beach looking at the water with your head tipped slightly sideways, your polarized sunglasses do not work very well. Why not?

Solution:

The axis of polarization for the sunglasses has been rotated 90°.

Problems

Exercise:

Problem:

What angle is needed between the direction of polarized light and the axis of a polarizing filter to cut its intensity in half?

Exercise:

Problem:

The angle between the axes of two polarizing filters is 45.0°. By how much does the second filter reduce the intensity of the light coming through the first?

Solution:

0.500

Exercise:

Problem:

Two polarizing sheets P_1 and P_2 are placed together with their transmission axes oriented at an angle θ to each other. What is θ when only 25% of the maximum transmitted light intensity passes through them?

Exercise:

Problem:

Suppose that in the preceding problem the light incident on P_1 is unpolarized. At the determined value of θ, what fraction of the incident light passes through the combination?

Solution:

0.125 or 1/8

Exercise:

Problem:

If you have completely polarized light of intensity $150\ \mathrm{W/m^2}$, what will its intensity be after passing through a polarizing filter with its axis at an 89.0° angle to the light's polarization direction?

Exercise:

Problem:

What angle would the axis of a polarizing filter need to make with the direction of polarized light of intensity $1.00\ \mathrm{kW/m^2}$ to reduce the intensity to $10.0\ \mathrm{W/m^2}$?

Solution:

84.3°

Exercise:

Problem:

At the end of [link], it was stated that the intensity of polarized light is reduced to 90.0% of its original value by passing through a polarizing filter with its axis at an angle of 18.4° to the direction of polarization. Verify this statement.

Exercise:

Problem:

Show that if you have three polarizing filters, with the second at an angle of 45.0° to the first and the third at an angle of 90.0° to the first, the intensity of light passed by the first will be reduced to 25.0% of its value. (This is in contrast to having only the first and third, which reduces the intensity to zero, so that placing the second between them increases the intensity of the transmitted light.)

Solution:

$0.250\ I_0$

Exercise:

Problem:

Three polarizing sheets are placed together such that the transmission axis of the second sheet is oriented at 25.0° to the axis of the first, whereas the transmission axis of the third sheet is oriented at 40.0° (in the same sense) to the axis of the first. What fraction of the intensity of an incident unpolarized beam is transmitted by the combination?

Exercise:

Problem:

In order to rotate the polarization axis of a beam of linearly polarized light by 90.0°, a student places sheets P_1 and P_2 with their transmission axes at 45.0° and 90.0°, respectively, to the beam's axis of polarization. (a) What fraction of the incident light passes through P_1 and (b) through the combination? (c) Repeat your calculations for part (b) for transmission-axis angles of 30.0° and 90.0°, respectively.

Solution:

a. 0.500; b. 0.250; c. 0.187

Exercise:

Problem:

It is found that when light traveling in water falls on a plastic block, Brewster's angle is 50.0°. What is the refractive index of the plastic?

Exercise:

Problem:

At what angle will light reflected from diamond be completely polarized?

Solution:

67.54°

Exercise:

Problem:

What is Brewster’s angle for light traveling in water that is reflected from crown glass?

Exercise:

Problem:

A scuba diver sees light reflected from the water’s surface. At what angle relative to the water’s surface will this light be completely polarized?

Solution:

53.1°

Additional Problems

Exercise:

Problem:

From his measurements, Roemer estimated that it took 22 min for light to travel a distance equal to the diameter of Earth’s orbit around the Sun. (a) Use this estimate along with the known diameter of Earth’s orbit to obtain a rough value of the speed of light. (b) Light actually takes 16.5 min to travel this distance. Use this time to calculate the speed of light.

Exercise:

Problem:

Cornu performed Fizeau’s measurement of the speed of light using a wheel of diameter 4.00 cm that contained 180 teeth. The distance from the wheel to the mirror was 22.9 km. Assuming he measured the speed of light accurately, what was the angular velocity of the wheel?

Solution:

114 radian/s

Exercise:

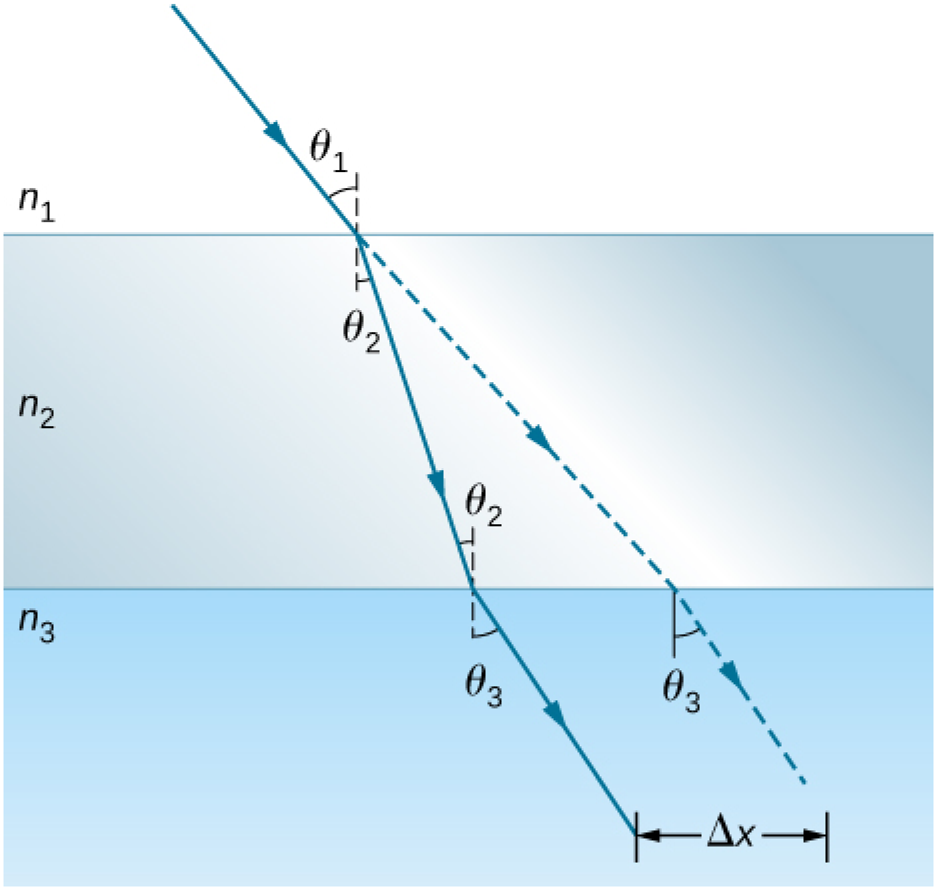
Problem:

Suppose you have an unknown clear substance immersed in water, and you wish to identify it by finding its index of refraction. You arrange to have a beam of light enter it at an angle of 45.0°, and you observe the angle of refraction to be 40.3°. What is the index of refraction of the substance and its likely identity?

Exercise:

Problem:

Shown below is a ray of light going from air through crown glass into water, such as going into a fish tank. Calculate the amount the ray is displaced by the glass (Δx), given that the incident angle is 40.0° and the glass is 1.00 cm thick.



Solution:

3.72 mm

Exercise:

Problem:

Considering the previous problem, show that θ_3 is the same as it would be if the second medium were not present.

Exercise:

Problem:

At what angle is light inside crown glass completely polarized when reflected from water, as in a fish tank?

Solution:

41.2°

Exercise:

Problem:

Light reflected at 55.6° from a window is completely polarized. What is the window’s index of refraction and the likely substance of which it is made?

Exercise:

Problem:

(a) Light reflected at 62.5° from a gemstone in a ring is completely polarized. Can the gem be a diamond? (b) At what angle would the light be completely polarized if the gem was in water?

Solution:

a. 1.92. The gem is not a diamond (it is zircon). b. 55.2°

Exercise:

Problem:

If θ_{b} is Brewster’s angle for light reflected from the top of an interface between two substances, and θ'_{b} is Brewster’s angle for light reflected from below, prove that $\theta_{\text{b}} + \theta'_{\text{b}} = 90.0^\circ$.

Exercise:

Problem:

Unreasonable results Suppose light travels from water to another substance, with an angle of incidence of 10.0° and an angle of refraction of 14.9°. (a) What is the index of refraction of the other substance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. 0.898; b. We cannot have $n < 1.00$, since this would imply a speed greater than c. c. The refracted angle is too big relative to the angle of incidence.

Exercise:

Problem:

Unreasonable results Light traveling from water to a gemstone strikes the surface at an angle of 80.0° and has an angle of refraction of 15.2° . (a) What is the speed of light in the gemstone? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem:

If a polarizing filter reduces the intensity of polarized light to 50.0% of its original value, by how much are the electric and magnetic fields reduced?

Solution:

$0.707\ B_1$

Exercise:

Problem:

Suppose you put on two pairs of polarizing sunglasses with their axes at an angle of 15.0°. How much longer will it take the light to deposit a given amount of energy in your eye compared with a single pair of sunglasses? Assume the lenses are clear except for their polarizing characteristics.

Exercise:

Problem:

(a) On a day when the intensity of sunlight is $1.00\,\text{kW/m}^2$, a circular lens 0.200 m in diameter focuses light onto water in a black beaker. Two polarizing sheets of plastic are placed in front of the lens with their axes at an angle of 20.0°. Assuming the sunlight is unpolarized and the polarizers are 100% efficient, what is the initial rate of heating of the water in $^\circ\text{C/s}$, assuming it is 80.0% absorbed? The aluminum beaker has a mass of 30.0 grams and contains 250 grams of water. (b) Do the polarizing filters get hot? Explain.

Solution:

a. $1.69 \times 10^{-2}\,^\circ\text{C/s}$; b. yes

Challenge Problems

Exercise:

Problem:

Light shows staged with lasers use moving mirrors to swing beams and create colorful effects. Show that a light ray reflected from a mirror changes direction by 2θ when the mirror is rotated by an angle θ.

Exercise:

Problem:

Consider sunlight entering Earth's atmosphere at sunrise and sunset—that is, at a 90.0° incident angle. Taking the boundary between nearly empty space and the atmosphere to be sudden, calculate the angle of refraction for sunlight. This lengthens the time the Sun appears to be above the horizon, both at sunrise and sunset. Now construct a problem in which you determine the angle of refraction for different models of the atmosphere, such as various layers of varying density. Your instructor may wish to guide you on the level of complexity to consider and on how the index of refraction varies with air density.

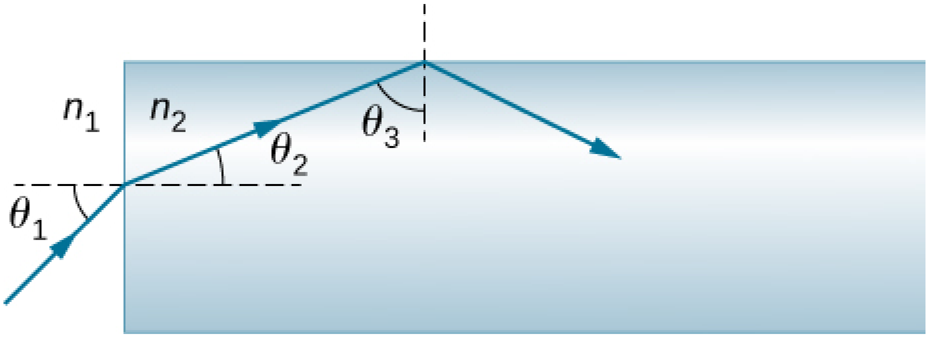
Solution:

First part: 88.6°. The remainder depends on the complexity of the solution the reader constructs.

Exercise:

Problem:

A light ray entering an optical fiber surrounded by air is first refracted and then reflected as shown below. Show that if the fiber is made from crown glass, any incident ray will be totally internally reflected.



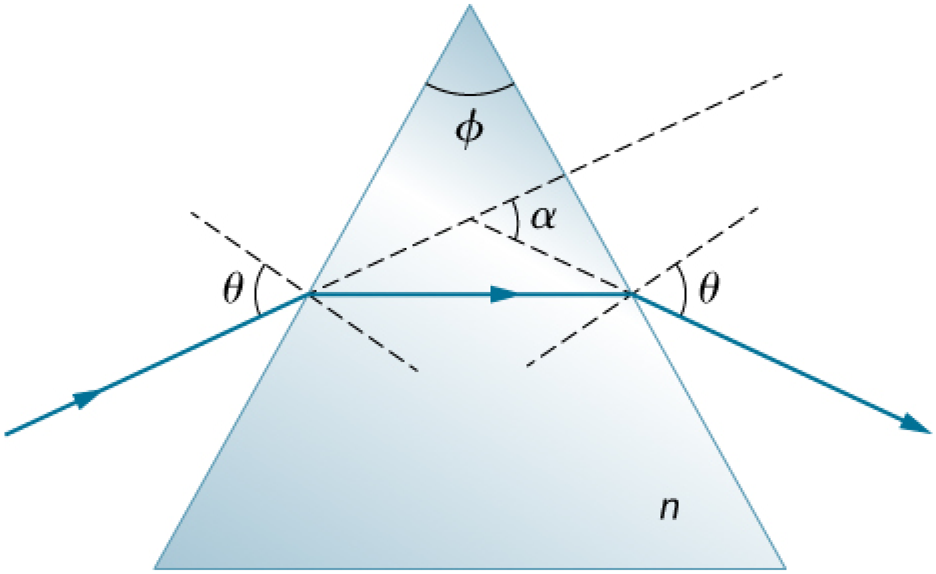
Exercise:

Problem:

A light ray falls on the left face of a prism (see below) at the angle of incidence θ for which the emerging beam has an angle of refraction θ at the right face. Show that the index of refraction n of the glass prism is given by

$$n = \frac{\sin \frac{1}{2}(\alpha + \phi)}{\sin \frac{1}{2}\phi}$$

where ϕ is the vertex angle of the prism and α is the angle through which the beam has been deviated. If $\alpha = 37.0^\circ$ and the base angles of the prism are each 50.0°, what is n?



Solution:

proof; 1.33

Exercise:

Problem:

If the apex angle ϕ in the previous problem is 20.0° and $n = 1.50$, what is the value of α?

Exercise:

Problem:

The light incident on polarizing sheet P_1 is linearly polarized at an angle of 30.0° with respect to the transmission axis of P_1. Sheet P_2 is placed so that its axis is parallel to the polarization axis of the incident light, that is, also at 30.0° with respect to P_1. (a) What fraction of the incident light passes through P_1? (b) What fraction of the incident light is passed by the combination? (c) By rotating P_2, a maximum in transmitted intensity is obtained. What is the ratio of this maximum intensity to the intensity of transmitted light when P_2 is at 30.0° with respect to P_1?

Solution:

a. 0.750; b. 0.563; c. 1.33

Exercise:

Problem:

Prove that if I is the intensity of light transmitted by two polarizing filters with axes at an angle θ and $I\prime$ is the intensity when the axes are at an angle $90.0^\circ - \theta$, then $I + I' = I_0$, the original intensity. (*Hint:* Use the trigonometric identities $\cos 90.0^\circ - \theta = \sin\theta$ and $\cos^2\theta + \sin^2\theta = 1.$)

Glossary

birefringent
refers to crystals that split an unpolarized beam of light into two beams

Brewster’s angle
angle of incidence at which the reflected light is completely polarized

Brewster’s law
$\tan\theta_\text{b} = \frac{n_2}{n_1}$, where n_1 is the medium in which the incident and reflected light travel and n_2 is the index of refraction of the medium that forms the interface that reflects the light

direction of polarization
direction parallel to the electric field for EM waves

horizontally polarized
oscillations are in a horizontal plane

Malus’s law
where I_0 is the intensity of the polarized wave before passing through the filter

optically active
substances that rotate the plane of polarization of light passing through them

polarization
attribute that wave oscillations have a definite direction relative to the direction of propagation of the wave

polarized
refers to waves having the electric and magnetic field oscillations in a definite direction

unpolarized
refers to waves that are randomly polarized

vertically polarized
oscillations are in a vertical plane

Introduction class="introduction"

Cloud Gate is a public sculpture by Anish Kapoor located in Millennium Park in Chicago. Its stainless steel plates reflect and distort images around it, including the Chicago skyline. Dedicated in 2006, it has become a popular tourist attraction, illustrating how art can use the principles of physical optics to startle and entertain. (credit: modificatio

n of work
by Dhilung
Kirat)

This chapter introduces the major ideas of geometric optics, which describe the formation of images due to reflection and refraction. It is called “geometric” optics because the images can be characterized using geometric constructions, such as ray diagrams. We have seen that visible light is an electromagnetic wave; however, its wave nature becomes evident only when light interacts with objects with dimensions comparable to the wavelength (about 500 nm for visible light). Therefore, the laws of geometric optics only apply to light interacting with objects much larger than the wavelength of the light.

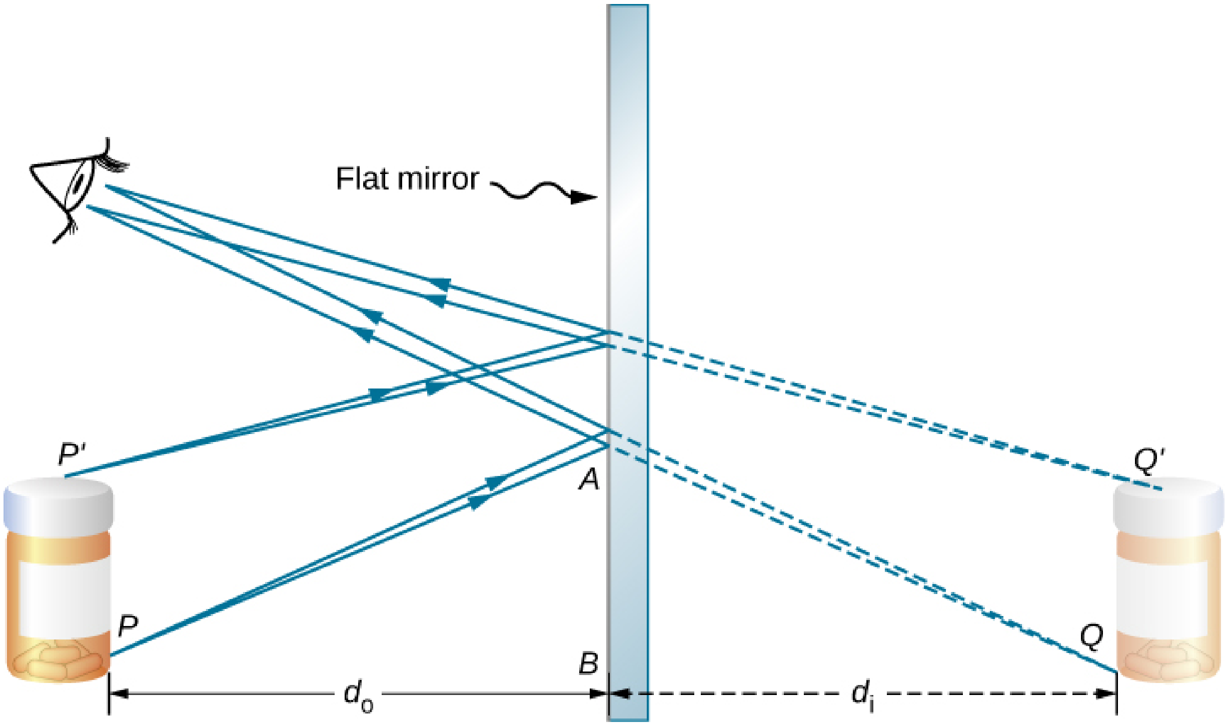
Images Formed by Plane Mirrors

By the end of this section, you will be able to:

- Describe how an image is formed by a plane mirror.
- Distinguish between real and virtual images.
- Find the location and characterize the orientation of an image created by a plane mirror.

You only have to look as far as the nearest bathroom to find an example of an image formed by a mirror. Images in a **plane mirror** are the same size as the object, are located behind the mirror, and are oriented in the same direction as the object (i.e., “upright”).

To understand how this happens, consider [link]. Two rays emerge from point P, strike the mirror, and reflect into the observer’s eye. Note that we use the law of reflection to construct the reflected rays. If the reflected rays are extended backward behind the mirror (see dashed lines in [link]), they seem to originate from point Q. This is where the image of point P is located. If we repeat this process for point P', we obtain its image at point Q'. You should convince yourself by using basic geometry that the image height (the distance from Q to Q') is the same as the object height (the distance from P to P'). By forming images of all points of the object, we obtain an upright image of the object behind the mirror.



Two light rays originating from point P on an object are reflected by a flat mirror into the eye of an observer. The reflected rays are obtained by using the law of reflection. Extending these reflected rays backward, they seem to come from point Q behind the mirror, which is where the virtual image is located. Repeating this process for point P' gives the image point Q'. The image height is thus the same as the object height, the image is upright, and the object distance d_o is the same as the image distance d_i. (credit: modification of work by Kevin Dufendach)

Notice that the reflected rays appear to the observer to come directly from the image behind the mirror. In reality, these rays come from the points on the mirror where they are reflected. The image behind the mirror is called a **virtual image** because it cannot be projected onto a screen—the rays only appear to originate from a common point behind the mirror. If you walk behind the mirror, you cannot see the image, because the rays do not go there. However, in front of the mirror, the rays behave exactly as if they come from behind the mirror, so that is where the virtual image is located.

Later in this chapter, we discuss real images; a **real image** can be projected onto a screen because the rays physically go through the image. You can certainly see both real and virtual images. The difference is that a virtual image cannot be projected onto a screen, whereas a real image can.

Locating an Image in a Plane Mirror

The law of reflection tells us that the angle of incidence is the same as the angle of reflection. Applying this to triangles *PAB* and *QAB* in [link] and using basic geometry shows that they are congruent triangles. This means that the distance *PB* from the object to the mirror is the same as the distance *BQ* from the mirror to the image. The **object distance** (denoted d_o) is the distance from the mirror to the object (or, more generally, from the center of the optical element that creates its image). Similarly, the **image distance** (denoted d_i) is the distance from the mirror to the image (or, more generally, from the center of the optical element that creates it). If we measure distances from the mirror, then the object and image are in opposite directions, so for a plane mirror, the object and image distances should have the opposite signs:

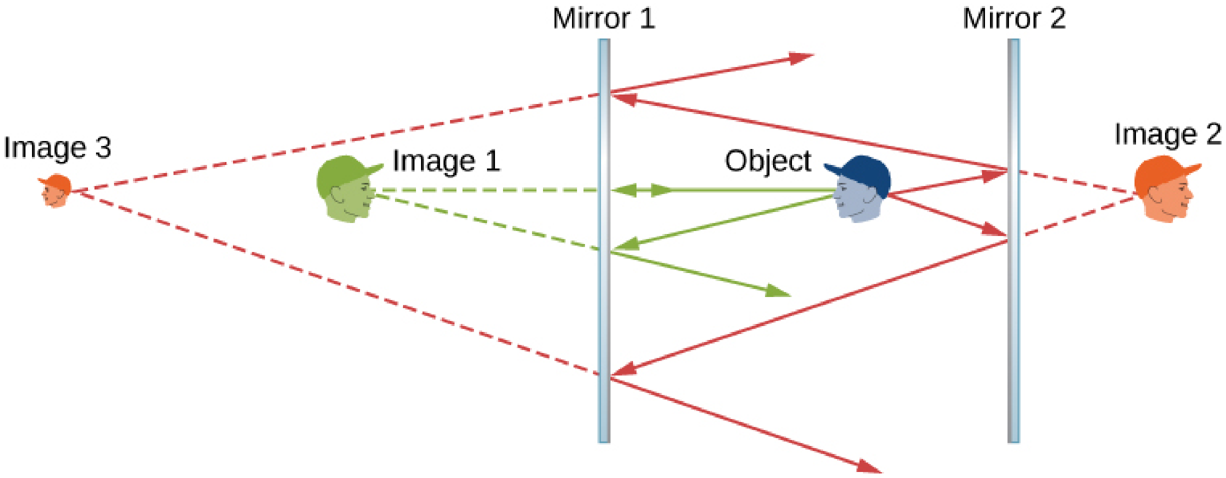
Note:
Equation:

$$d_{\text{o}} = -d_{\text{i}}.$$

An extended object such as the container in [link] can be treated as a collection of points, and we can apply the method above to locate the image of each point on the extended object, thus forming the extended image.

Multiple Images

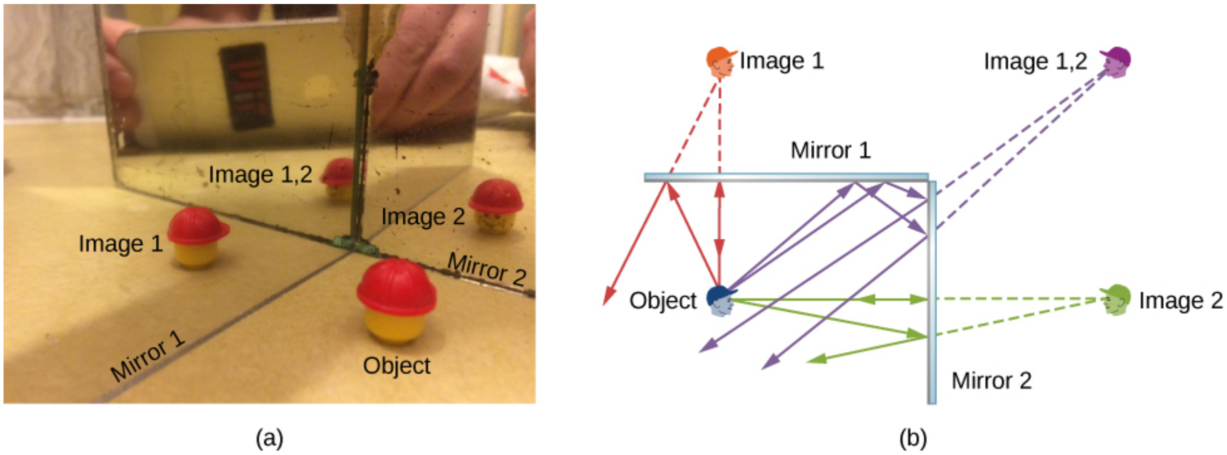
If an object is situated in front of two mirrors, you may see images in both mirrors. In addition, the image in the first mirror may act as an object for the second mirror, so the second mirror may form an image of the image. If the mirrors are placed parallel to each other and the object is placed at a point other than the midpoint between them, then this process of image-of-an-image continues without end, as you may have noticed when standing in a hallway with mirrors on each side. This is shown in [link], which shows three images produced by the blue object. Notice that each reflection reverses front and back, just like pulling a right-hand glove inside out produces a left-hand glove (this is why a reflection of your right hand is a left hand). Thus, the fronts and backs of images 1 and 2 are both inverted with respect to the object, and the front and back of image 3 is inverted with respect to image 2, which is the object for image 3.



Two parallel mirrors can produce, in theory, an infinite number of images of an object placed off center between the mirrors. Three of these images are shown here. The front and back of each image is inverted with respect to its object. Note that the colors are only to identify the images. For normal mirrors, the color of an image is essentially the same as that of its object.

You may have noticed that image 3 is smaller than the object, whereas images 1 and 2 are the same size as the object. The ratio of the image height with respect to the object height is called **magnification**. More will be said about magnification in the next section.

Infinite reflections may terminate. For instance, two mirrors at right angles form three images, as shown in part (a) of [link]. Images 1 and 2 result from rays that reflect from only a single mirror, but image 1,2 is formed by rays that reflect from both mirrors. This is shown in the ray-tracing diagram in part (b) of [link]. To find image 1,2, you have to look behind the corner of the two mirrors.



Two mirrors can produce multiple images. (a) Three images of a plastic head are visible in the two mirrors at a right angle. (b) A single object reflecting from two mirrors at a right angle can produce three images, as shown by the green, purple, and red images.

Summary

- A plane mirror always forms a virtual image (behind the mirror).
- The image and object are the same distance from a flat mirror, the image size is the same as the object size, and the image is upright.

Conceptual Questions

Exercise:

Problem:

What are the differences between real and virtual images? How can you tell (by looking) whether an image formed by a single lens or mirror is real or virtual?

Solution:

Virtual image cannot be projected on a screen. You cannot distinguish a real image from a virtual image simply by judging from the image perceived with your eye.

Exercise:

Problem: Can you see a virtual image? Explain your response.

Exercise:

Problem: Can you photograph a virtual image?

Solution:

Yes, you can photograph a virtual image. For example, if you photograph your reflection from a plane mirror, you get a photograph of a virtual image. The camera focuses the light that enters its lens to form an image; whether the source of the light is a real object or a reflection from mirror (i.e., a virtual image) does not matter.

Exercise:

Problem: Can you project a virtual image onto a screen?

Exercise:

Problem: Is it necessary to project a real image onto a screen to see it?

Solution:

No, you can see the real image the same way you can see the virtual image. The retina of your eye effectively serves as a screen.

Exercise:

Problem:

Devise an arrangement of mirrors allowing you to see the back of your head. What is the minimum number of mirrors needed for this task?

Exercise:

Problem:

If you wish to see your entire body in a flat mirror (from head to toe), how tall should the mirror be? Does its size depend upon your distance away from the mirror? Provide a sketch.

Solution:

The mirror should be half your size and its top edge should be at the level of your eyes. The size does not depend on your distance from the mirror.

Problems

Exercise:

Problem:

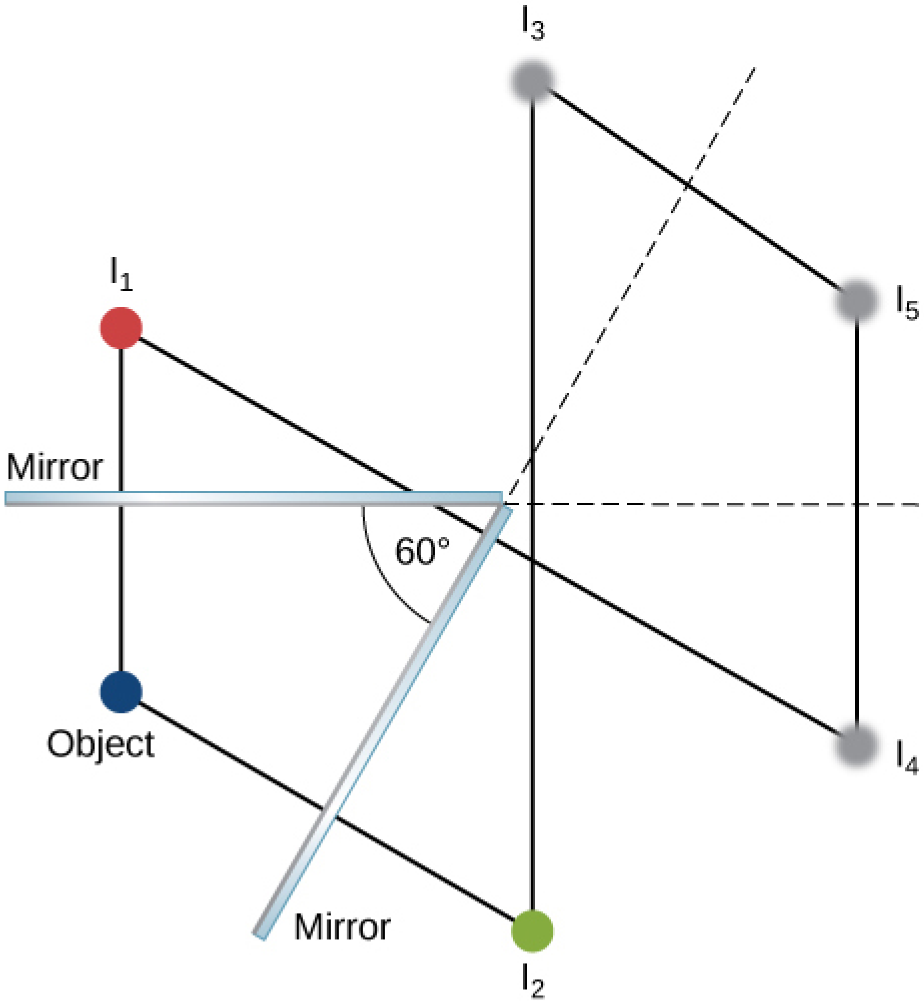
Consider a pair of flat mirrors that are positioned so that they form an angle of 120°. An object is placed on the bisector between the mirrors. Construct a ray diagram as in [link] to show how many images are formed.

Exercise:

Problem:

Consider a pair of flat mirrors that are positioned so that they form an angle of 60°. An object is placed on the bisector between the mirrors. Construct a ray diagram as in [link] to show how many images are formed.

Solution:



Exercise:

Problem:

By using more than one flat mirror, construct a ray diagram showing how to create an inverted image.

Glossary

plane mirror
plane (flat) reflecting surface

image distance
distance of the image from the central axis of the optical element that produces the image

magnification
ratio of image size to object size

object distance
distance of the object from the central axis of the optical element that produces its image

real image
image that can be projected onto a screen because the rays physically go through the image

virtual image
image that cannot be projected on a screen because the rays do not physically go through the image, they only appear to originate from the image

Spherical Mirrors

By the end of this section, you will be able to:

- Describe image formation by spherical mirrors.
- Use ray diagrams and the mirror equation to calculate the properties of an image in a spherical mirror.

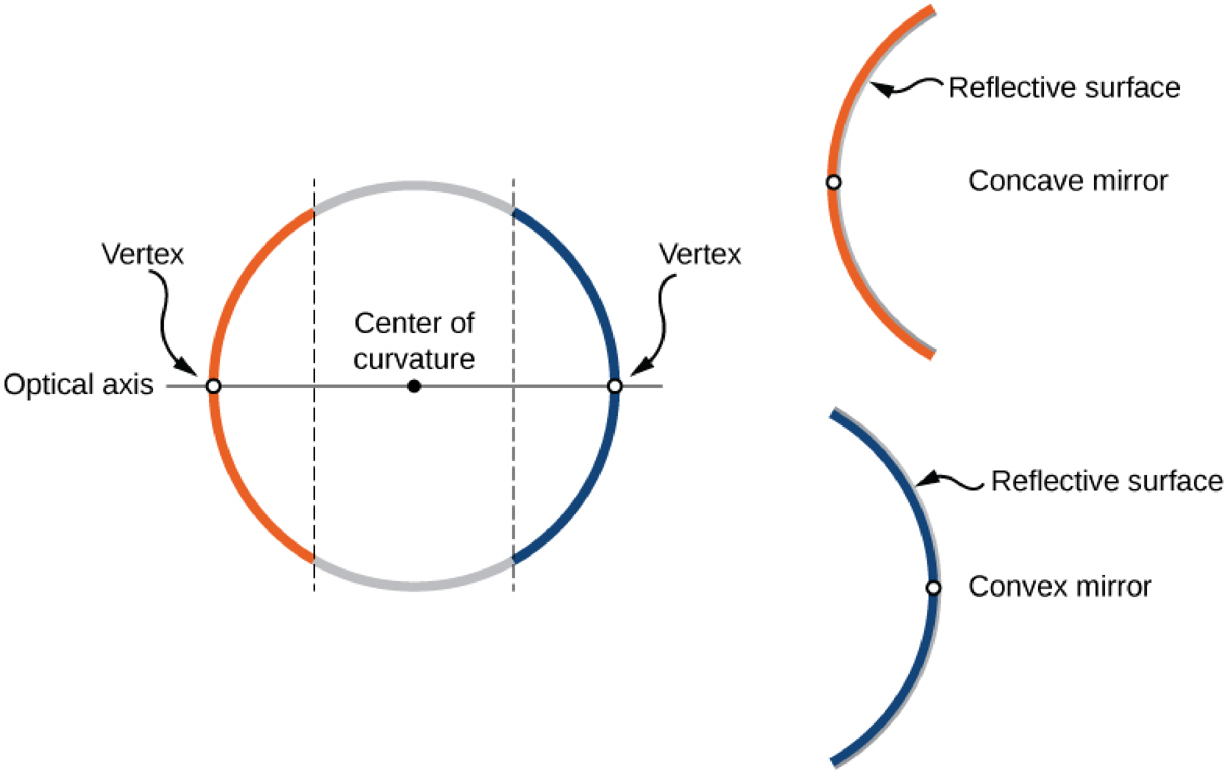
The image in a plane mirror has the same size as the object, is upright, and is the same distance behind the mirror as the object is in front of the mirror. A **curved mirror**, on the other hand, can form images that may be larger or smaller than the object and may form either in front of the mirror or behind it. In general, any curved surface will form an image, although some images make be so distorted as to be unrecognizable (think of fun house mirrors).

Because curved mirrors can create such a rich variety of images, they are used in many optical devices that find many uses. We will concentrate on spherical mirrors for the most part, because they are easier to manufacture than mirrors such as parabolic mirrors and so are more common.

Curved Mirrors

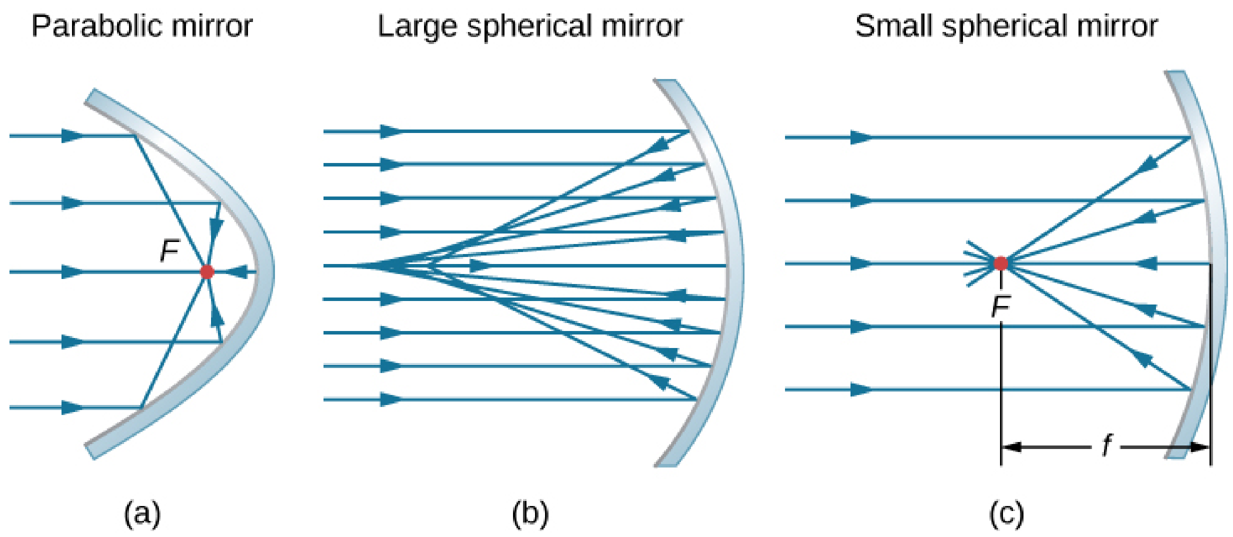
We can define two general types of spherical mirrors. If the reflecting surface is the outer side of the sphere, the mirror is called a **convex mirror**. If the inside surface is the reflecting surface, it is called a **concave mirror**.

Symmetry is one of the major hallmarks of many optical devices, including mirrors and lenses. The symmetry axis of such optical elements is often called the principal axis or **optical axis**. For a spherical mirror, the optical axis passes through the mirror’s center of curvature and the mirror’s **vertex**, as shown in [link].



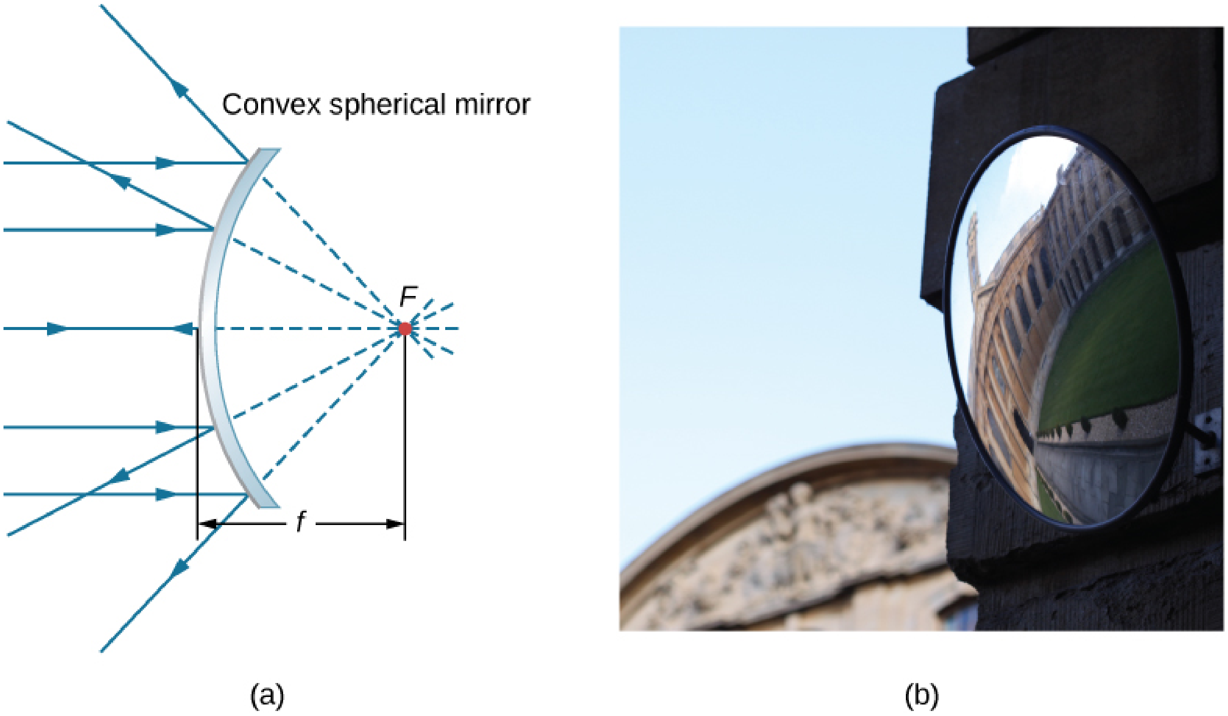
A spherical mirror is formed by cutting out a piece of a sphere and silvering either the inside or outside surface. A concave mirror has silvering on the interior surface (think “cave”), and a convex mirror has silvering on the exterior surface.

Consider rays that are parallel to the optical axis of a parabolic mirror, as shown in part (a) of [link]. Following the law of reflection, these rays are reflected so that they converge at a point, called the **focal point**. Part (b) of this figure shows a spherical mirror that is large compared with its radius of curvature. For this mirror, the reflected rays do not cross at the same point, so the mirror does not have a well-defined focal point. This is called spherical aberration and results in a blurred image of an extended object. Part (c) shows a spherical mirror that is small compared to its radius of curvature. This mirror is a good approximation of a parabolic mirror, so rays that arrive parallel to the optical axis are reflected to a well-defined focal point. The distance along the optical axis from the mirror to the focal point is called the **focal length** of the mirror.



(a) Parallel rays reflected from a parabolic mirror cross at a single point called the focal point *F*. (b) Parallel rays reflected from a large spherical mirror do not cross at a common point. (c) If a spherical mirror is small compared with its radius of curvature, it better approximates the central part of a parabolic mirror, so parallel rays essentially cross at a common point. The distance along the optical axis from the mirror to the focal point is the focal length *f* of the mirror.

A convex spherical mirror also has a focal point, as shown in [link]. Incident rays parallel to the optical axis are reflected from the mirror and seem to originate from point *F* at focal length *f* behind the mirror. Thus, the focal point is virtual because no real rays actually pass through it; they only appear to originate from it.

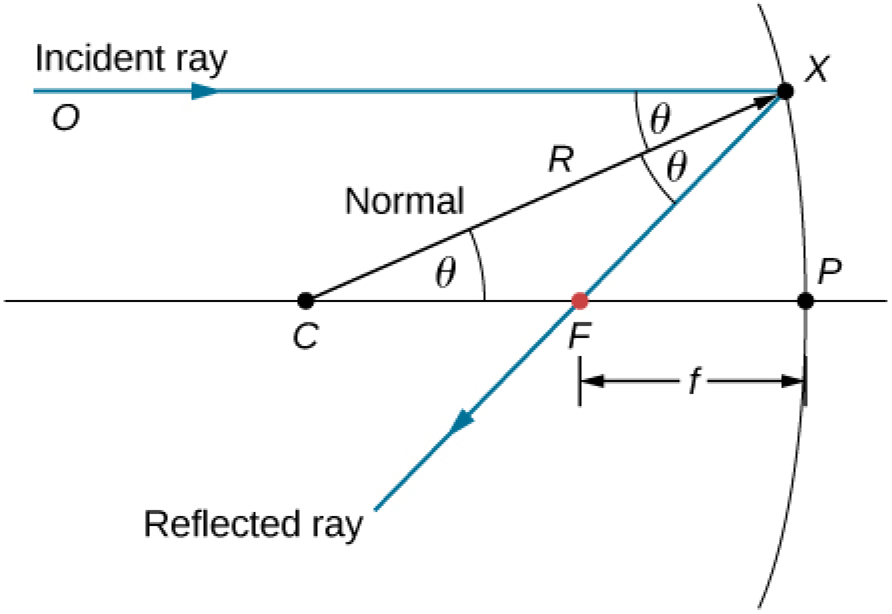


(a) Rays reflected by a convex spherical mirror: Incident rays of light parallel to the optical axis are reflected from a convex spherical mirror and seem to originate from a well-defined focal point at focal distance *f* on the opposite side of the mirror. The focal point is virtual because no real rays pass through it. (b) Photograph of a virtual image formed by a convex mirror. (credit b: modification of work by Jenny Downing)

How does the focal length of a mirror relate to the mirror's radius of curvature? [link] shows a single ray that is reflected by a spherical concave mirror. The incident ray is parallel to the optical axis. The point at which the reflected ray crosses the optical axis is the focal point. Note that all incident rays that are parallel to the optical axis are reflected through the focal point—we only show one ray for simplicity. We want to find how the focal length *FP* (denoted by *f*) relates to the radius of curvature of the mirror, *R*, whose length is $R = CF + FP$. The law of reflection tells us that angles *OXC* and *CXF* are the same, and because the incident ray is parallel to the optical axis, angles *OXC* and *XCP* are also the same. Thus, triangle *CXF* is an isosceles triangle with $CF = FX$. If the angle θ is small (so that $\sin \theta \approx \theta$; this is called the “small-angle approximation”), then $FX \approx FP$ or $CF \approx FP$. Inserting this into the equation for the radius *R*, we get

Equation:

$$R = CF + FP = FP + FP = 2FP = 2f$$



Reflection in a concave mirror. In the small-angle approximation, a ray that is parallel to the optical axis *CP* is reflected through the focal point *F* of the mirror.

In other words, in the small-angle approximation, the focal length *f* of a concave spherical mirror is half of its radius of curvature, *R*:

Note:
Equation:

$$f = \frac{R}{2}.$$

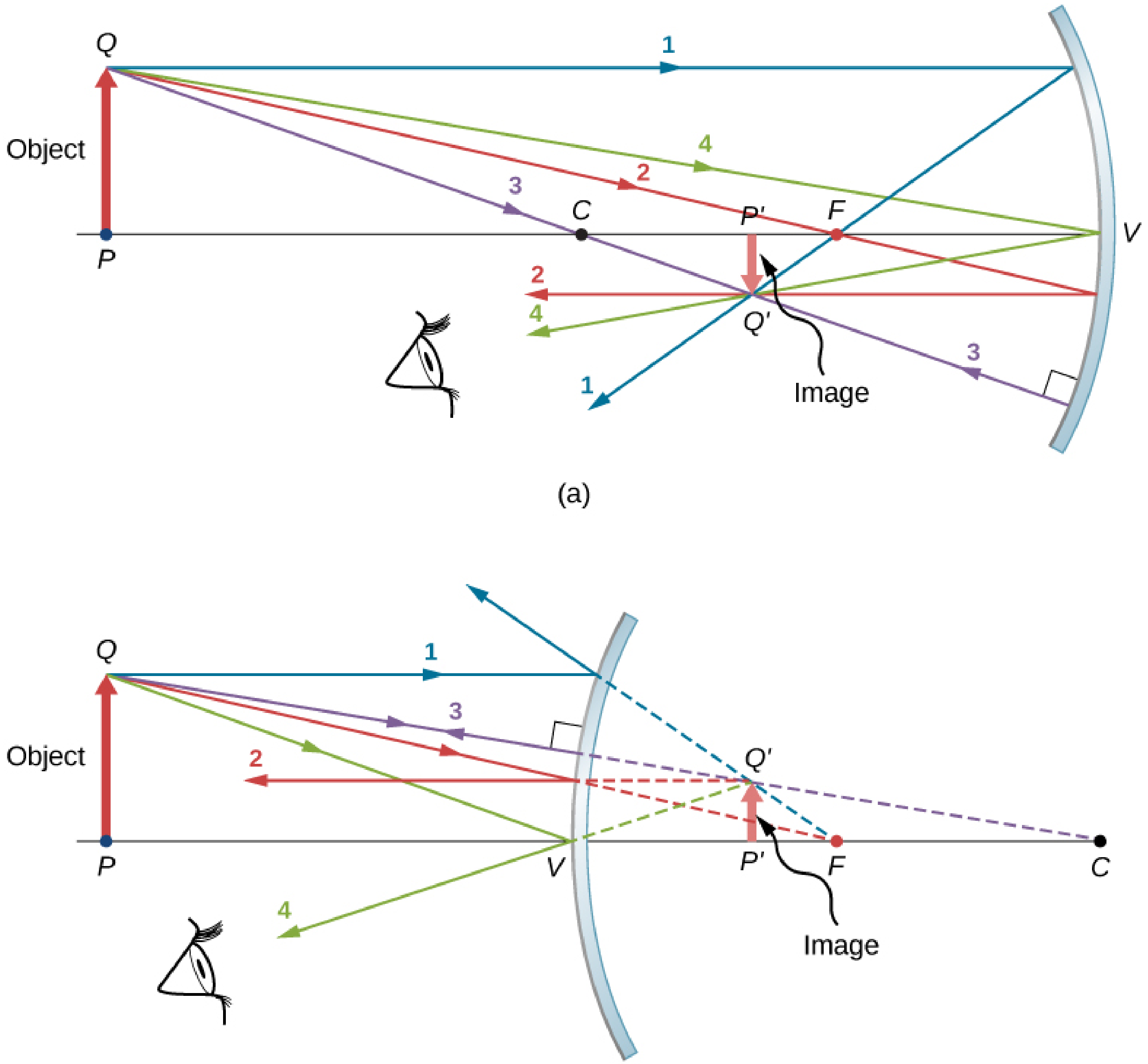
In this chapter, we assume that the **small-angle approximation** (also called the paraxial approximation) is always valid. In this approximation, all rays are paraxial rays, which means that they make a small angle with the optical axis and are at a distance much less than the radius of curvature from the optical axis. In this case, their angles θ of reflection are small angles, so $\sin\theta \approx \tan\theta \approx \theta$.

Using Ray Tracing to Locate Images

To find the location of an image formed by a spherical mirror, we first use ray tracing, which is the technique of drawing rays and using the law of reflection to determine the reflected rays (later, for lenses, we use the law of refraction to determine refracted rays). Combined with some basic geometry, we can use ray tracing to find the focal point, the image location, and other information about how a mirror manipulates light. In fact, we already used ray tracing above to locate the focal point of spherical mirrors, or the image distance of flat mirrors. To locate the image of an object, you must locate at least two points of the image. Locating each point requires drawing at least two rays from a point on the object and constructing their reflected rays. The point at which the reflected rays intersect, either in real space or in virtual space, is where the corresponding point of the image is located. To make ray tracing easier, we concentrate on four “principal” rays whose reflections are easy to construct.

[link] shows a concave mirror and a convex mirror, each with an arrow-shaped object in front of it. These are the objects whose images we want to locate by ray tracing. To do so, we draw rays from point Q that is on the object but not on the optical axis. We choose to draw our ray from the tip of the object. Principal ray 1 goes from point Q and travels parallel to the optical axis. The reflection of this ray must pass through the focal point, as discussed above. Thus, for the concave mirror, the reflection of principal

ray 1 goes through focal point *F*, as shown in part (b) of the figure. For the convex mirror, the backward extension of the reflection of principal ray 1 goes through the focal point (i.e., a virtual focus). Principal ray 2 travels first on the line going through the focal point and then is reflected back along a line parallel to the optical axis. Principal ray 3 travels toward the center of curvature of the mirror, so it strikes the mirror at normal incidence and is reflected back along the line from which it came. Finally, principal ray 4 strikes the vertex of the mirror and is reflected symmetrically about the optical axis.



(a)

(b)

The four principal rays shown for both (a) a concave mirror and (b) a convex mirror. The image forms where the rays intersect (for real images) or where their backward extensions intersect (for virtual images).

The four principal rays intersect at point Q', which is where the image of point Q is located. To locate point Q', drawing any two of these principle rays would suffice. We are thus free to choose whichever of the principal rays we desire to locate the image. Drawing more than two principal rays is sometimes useful to verify that the ray tracing is correct.

To completely locate the extended image, we need to locate a second point in the image, so that we know how the image is oriented. To do this, we trace the principal rays from the base of the object. In this case, all four principal rays run along the optical axis, reflect from the mirror, and then run back along the optical axis. The difficulty is that, because these rays are collinear, we cannot determine a unique point where they intersect. All we know is that the base of the image is on the optical axis. However, because the mirror is symmetrical from top to bottom, it does not change the vertical orientation of the object. Thus, because the object is vertical, the image must be vertical. Therefore, the image of the base of the object is on the optical axis directly above the image of the tip, as drawn in the figure.

For the concave mirror, the extended image in this case forms between the focal point and the center of curvature of the mirror. It is inverted with respect to the object, is a real image, and is smaller than the object. Were we to move the object closer to or farther from the mirror, the characteristics of the image would change. For example, we show, as a later exercise, that an object placed between a concave mirror and its focal point leads to a virtual image that is upright and larger than the object. For the convex mirror, the extended image forms between the focal point and the mirror. It is upright with respect to the object, is a virtual image, and is smaller than the object.

Summary of Ray-Tracing Rules

Ray tracing is very useful for mirrors. The rules for ray tracing are summarized here for reference:

- A ray travelling parallel to the optical axis of a spherical mirror is reflected along a line that goes through the focal point of the mirror (ray 1 in [link]).
- A ray travelling along a line that goes through the focal point of a spherical mirror is reflected along a line parallel to the optical axis of the mirror (ray 2 in [link]).

- A ray travelling along a line that goes through the center of curvature of a spherical mirror is reflected back along the same line (ray 3 in [link]).
- A ray that strikes the vertex of a spherical mirror is reflected symmetrically about the optical axis of the mirror (ray 4 in [link]).

We use ray tracing to illustrate how images are formed by mirrors and to obtain numerical information about optical properties of the mirror. If we assume that a mirror is small compared with its radius of curvature, we can also use algebra and geometry to derive a mirror equation, which we do in the next section. Combining ray tracing with the mirror equation is a good way to analyze mirror systems.

Image Formation by Reflection—The Mirror Equation

For a plane mirror, we showed that the image formed has the same height and orientation as the object, and it is located at the same distance behind the mirror as the object is in front of the mirror. Although the situation is a bit more complicated for curved mirrors, using geometry leads to simple formulas relating the object and image distances to the focal lengths of concave and convex mirrors.

Consider the object *OP* shown in [link]. The center of curvature of the mirror is labeled *C* and is a distance *R* from the vertex of the mirror, as marked in the figure. The object and image distances are labeled d_o and d_i, and the object and image heights are labeled h_o and h_i, respectively. Because the angles ϕ and ϕ' are alternate interior angles, we know that they have the same magnitude. However, they must differ in sign if we measure angles from the optical axis, so $\phi = -\phi'$. An analogous scenario holds for the angles θ and θ'. The law of reflection tells us that they have the same magnitude, but their signs must differ if we measure angles from the optical axis. Thus, $\theta = -\theta'$. Taking the tangent of the angles θ and θ', and using the property that $\tan(-\theta) = -\tan\theta$, gives us

Equation:

$$\left.\begin{array}{l}\tan\theta = \frac{h_{\text{o}}}{d_{\text{o}}}\\ \tan\theta' = -\tan\theta = \frac{h_{\text{i}}}{d_{\text{i}}}\end{array}\right\} \frac{h_{\text{o}}}{d_{\text{o}}} = -\frac{h_{\text{i}}}{d_{\text{i}}} \text{ or } -\frac{h_{\text{o}}}{h_{\text{i}}} = \frac{d_{\text{o}}}{d_{\text{i}}}.$$

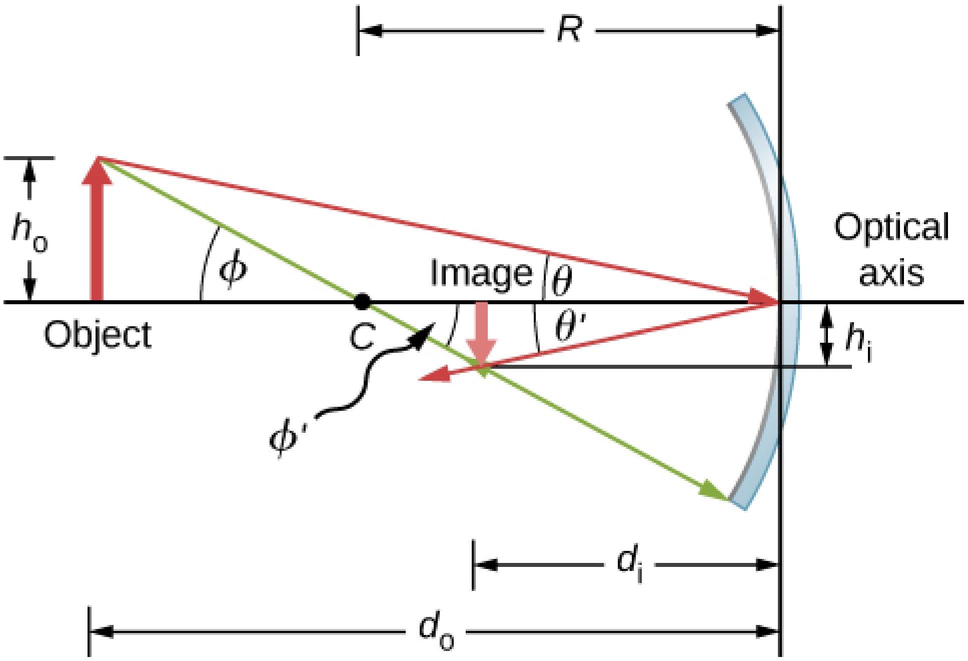


Image formed by a concave mirror.

Similarly, taking the tangent of ϕ and ϕ' gives

Equation:

$$\left.\begin{array}{l}\tan\phi = \frac{h_{\text{o}}}{d_{\text{o}}-R}\\ \tan\phi' = -\tan\phi = \frac{h_{\text{i}}}{R-d_{\text{i}}}\end{array}\right\} \frac{h_{\text{o}}}{d_{\text{o}}-R} = -\frac{h_{\text{i}}}{R-d_{\text{i}}} \text{ or } -\frac{h_{\text{o}}}{h_{\text{i}}} = \frac{d_{\text{o}}-R}{R-d_{\text{i}}}.$$

Combining these two results gives

Equation:

$$\frac{d_o}{d_i} = \frac{d_o - R}{R - d_i}.$$

After a little algebra, this becomes
Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{2}{R}.$$

No approximation is required for this result, so it is exact. However, as discussed above, in the small-angle approximation, the focal length of a spherical mirror is one-half the radius of curvature of the mirror, or $f = R/2$. Inserting this into [link] gives the *mirror equation*:

Note:
Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

The mirror equation relates the image and object distances to the focal distance and is valid only in the small-angle approximation. Although it was derived for a concave mirror, it also holds for convex mirrors (proving this is left as an exercise). We can extend the mirror equation to the case of a plane mirror by noting that a plane mirror has an infinite radius of curvature. This means the focal point is at infinity, so the mirror equation simplifies to
Equation:

$$d_o = -d_i$$

which is the same as [link] obtained earlier.

Notice that we have been very careful with the signs in deriving the mirror equation. For a plane mirror, the image distance has the opposite sign of the object distance. Also, the real image formed by the concave mirror in [link] is on the opposite side of the optical axis with respect to the object. In this case, the image height should have the opposite sign of the object height. To keep track of the signs of the various quantities in the mirror equation, we now introduce a sign convention.

Sign convention for spherical mirrors

Using a consistent sign convention is very important in geometric optics. It assigns positive or negative values for the quantities that characterize an optical system. Understanding the sign convention allows you to describe an image without constructing a ray diagram. This text uses the following sign convention:

1. The focal length f is positive for concave mirrors and negative for convex mirrors.
2. The image distance d_i is positive for real images and negative for virtual images.

Notice that rule 1 means that the radius of curvature of a spherical mirror can be positive or negative. What does it mean to have a negative radius of curvature? This means simply that the radius of curvature for a convex mirror is defined to be negative.

Image magnification

Let’s use the sign convention to further interpret the derivation of the mirror equation. In deriving this equation, we found that the object and image heights are related by

Equation:

$$-\frac{h_{\text{o}}}{h_{\text{i}}} = \frac{d_{\text{o}}}{d_{\text{i}}}.$$

See [link]. Both the object and the image formed by the mirror in [link] are real, so the object and image distances are both positive. The highest point of the object is above the optical axis, so the object height is positive. The image, however, is below the optical axis, so the image height is negative. Thus, this sign convention is consistent with our derivation of the mirror equation.

[link] in fact describes the **linear magnification** (often simply called "magnification") of the image in terms of the object and image distances. We thus define the dimensionless magnification *m* as follows:

Equation:

$$m = \frac{h_{\text{i}}}{h_{\text{o}}}.$$

If *m* is positive, the image is upright, and if *m* is negative, the image is inverted. If $|m| > 1$, the image is larger than the object, and if $|m| < 1$, the image is smaller than the object. With this definition of magnification, we get the following relation between the vertical and horizontal object and image distances:

Note:

Equation:

$$m = \frac{h_{\text{i}}}{h_{\text{o}}} = -\frac{d_{\text{i}}}{d_{\text{o}}}.$$

This is a very useful relation because it lets you obtain the magnification of the image from the object and image distances, which you can obtain from the mirror equation.

Example:

Solar Electric Generating System

One of the solar technologies used today for generating electricity involves a device (called a parabolic trough or concentrating collector) that concentrates sunlight onto a blackened pipe that contains a fluid. This heated fluid is pumped to a heat exchanger, where the thermal energy is transferred to another system that is used to generate steam and eventually generates electricity through a conventional steam cycle. [link] shows such a working system in southern California. The real mirror is a parabolic cylinder with its focus located at the pipe; however, we can approximate the mirror as exactly one-quarter of a circular cylinder.

Parabolic trough collectors are used to generate electricity in southern California. (credit: “kjkolb”/Wikimedia Commons)

a. If we want the rays from the sun to focus at 40.0 cm from the mirror, what is the radius of the mirror?
b. What is the amount of sunlight concentrated onto the pipe, per meter of pipe length, assuming the insolation (incident solar radiation) is $900\ \text{W/m}^2$?

c. If the fluid-carrying pipe has a 2.00-cm diameter, what is the temperature increase of the fluid per meter of pipe over a period of 1 minute? Assume that all solar radiation incident on the reflector is absorbed by the pipe, and that the fluid is mineral oil.

Strategy

First identify the physical principles involved. Part (a) is related to the optics of spherical mirrors. Part (b) involves a little math, primarily geometry. Part (c) requires an understanding of heat and density.

Solution

a. The sun is the object, so the object distance is essentially infinity: $d_o = \infty$. The desired image distance is $d_i = 40.0\text{ cm}$. We use the mirror equation to find the focal length of the mirror:

Equation:

$$\begin{aligned} \frac{1}{d_o} + \frac{1}{d_i} &= \frac{1}{f} \\ f &= \left(\frac{1}{d_o} + \frac{1}{d_i}\right)^{-1} \\ &= \left(\frac{1}{\infty} + \frac{1}{40.0\text{ cm}}\right)^{-1} \\ &= 40.0\text{ cm} \end{aligned}$$

Thus, the radius of the mirror is $R = 2f = 80.0\text{ cm}$.

b. The insolation is 900 W/m^2. You must find the cross-sectional area A of the concave mirror, since the power delivered is $900\text{ W/m}^2 \times A$. The mirror in this case is estimated as a quarter-section of a cylinder, so the area for a length L of the mirror is $A = \frac{1}{4}(2\pi R)L$. The area for a length of 1.00 m is then

Equation:

$$A = \frac{\pi}{2}R(1.00\text{ m}) = \frac{(3.14)}{2}(0.800\text{ m})(1.00\text{ m}) = 1.26\text{ m}^2.$$

The insolation on the 1.00-m length of pipe is then

Equation:

$$\left(9.00\ \times\ 10^2 \frac{\text{W}}{\text{m}^2}\right)\left(1.26\,\text{m}^2\right) = 1130\,\text{W}.$$

c. The increase in temperature is given by $Q = mc\Delta T$. The mass *m* of the mineral oil in the one-meter section of pipe is
Equation:

$$\begin{aligned} m &= \rho V = \rho\pi\left(\tfrac{d}{2}\right)^2(1.00\,\text{m}) \\ &= \left(8.00\ \times\ 10^2\,\text{kg/m}^3\right)(3.14)(0.0100\,\text{m})^2(1.00\,\text{m}) \\ &= 0.251\,\text{kg} \end{aligned}$$

Therefore, the increase in temperature in one minute is
Equation:

$$\begin{aligned} \Delta T &= Q/mc \\ &= \frac{(1130\,\text{W})(60.0\,\text{s})}{(0.251\,\text{kg})(1670\,\text{J·kg/}^\circ\text{C})} \\ &= 162^\circ\text{C} \end{aligned}$$

Significance
An array of such pipes in the California desert can provide a thermal output of 250 MW on a sunny day, with fluids reaching temperatures as high as 400°C. We are considering only one meter of pipe here and ignoring heat losses along the pipe.

Example:
Image in a Convex Mirror
A keratometer is a device used to measure the curvature of the cornea of the eye, particularly for fitting contact lenses. Light is reflected from the cornea, which acts like a convex mirror, and the keratometer measures the

magnification of the image. The smaller the magnification, the smaller the radius of curvature of the cornea. If the light source is 12 cm from the cornea and the image magnification is 0.032, what is the radius of curvature of the cornea?

Strategy

If you find the focal length of the convex mirror formed by the cornea, then you know its radius of curvature (it’s twice the focal length). The object distance is $d_\text{o} = 12\,\text{cm}$ and the magnification is $m = 0.032$. First find the image distance d_i and then solve for the focal length f.

Solution

Start with the equation for magnification, $m = -d_\text{i}/d_\text{o}$. Solving for d_i and inserting the given values yields

Equation:

$$d_\text{i} = -md_\text{o} = -(0.032)(12\,\text{cm}) = -0.384\,\text{cm}$$

where we retained an extra significant figure because this is an intermediate step in the calculation. Solve the mirror equation for the focal length f and insert the known values for the object and image distances. The result is

Equation:

$$\begin{aligned} \frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} &= \frac{1}{f} \\ f &= \left(\frac{1}{d_\text{o}} + \frac{1}{d_\text{i}}\right)^{-1} \\ &= \left(\frac{1}{12\,\text{cm}} + \frac{1}{-0.384\,\text{cm}}\right)^{-1} \\ &= -0.40\,\text{cm} \end{aligned}$$

The radius of curvature is twice the focal length, so

Equation:

$$R = 2f = -0.80\,\text{cm}$$

Significance

The focal length is negative, so the focus is virtual, as expected for a concave mirror and a real object. The radius of curvature found here is reasonable for a cornea. The distance from cornea to retina in an adult eye is about 2.0 cm. In practice, corneas may not be spherical, which complicates the job of fitting contact lenses. Note that the image distance here is negative, consistent with the fact that the image is behind the mirror. Thus, the image is virtual because no rays actually pass through it. In the problems and exercises, you will show that, for a fixed object distance, a smaller radius of curvature corresponds to a smaller the magnification.

Note:

Problem-Solving Strategy: Spherical Mirrors

Step 1. First make sure that image formation by a spherical mirror is involved.

Step 2. Determine whether ray tracing, the mirror equation, or both are required. A sketch is very useful even if ray tracing is not specifically required by the problem. Write symbols and known values on the sketch.

Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).

Step 4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns).

Step 5. If ray tracing is required, use the ray-tracing rules listed near the beginning of this section.

Step 6. Most quantitative problems require using the mirror equation. Use the examples as guides for using the mirror equation.

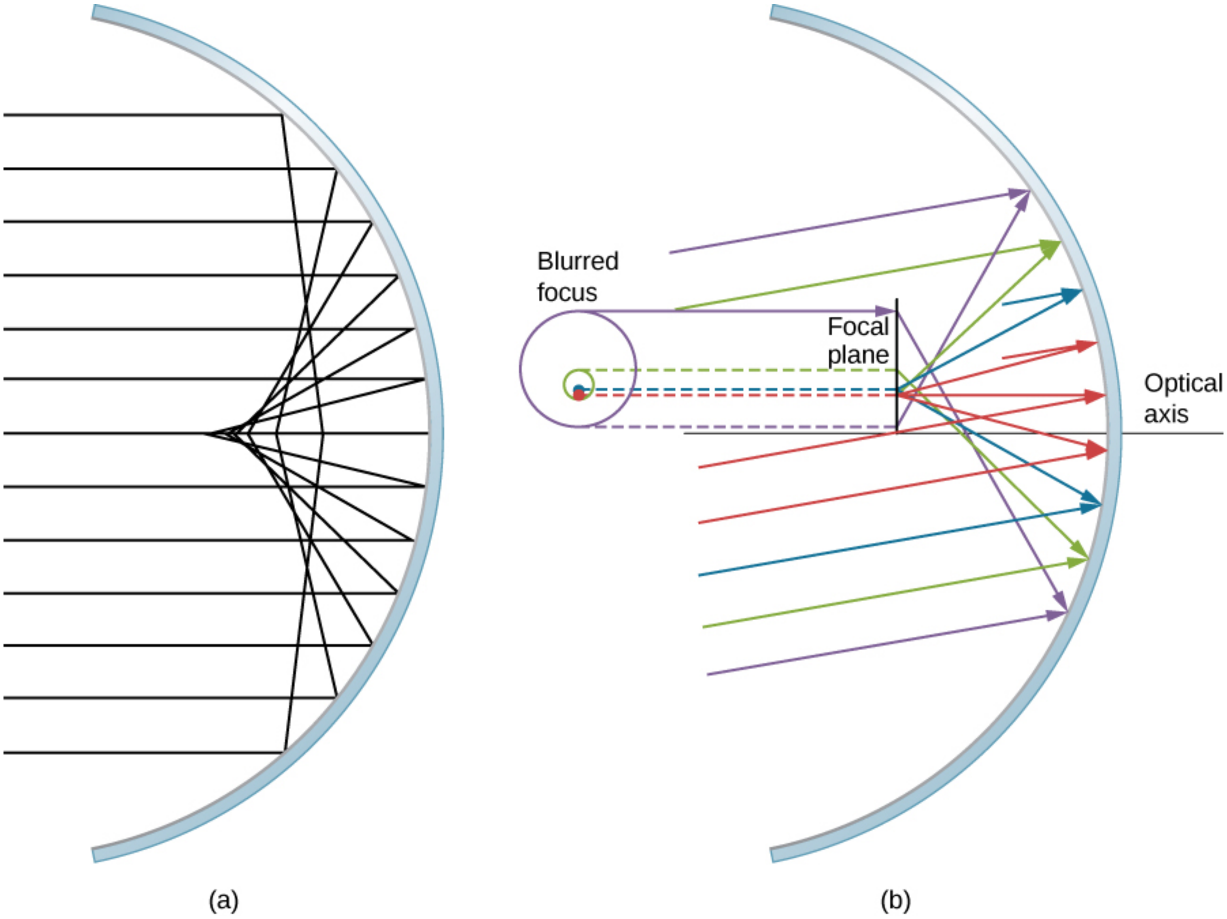
Step 7. Check to see whether the answer makes sense. Do the signs of object distance, image distance, and focal length correspond with what is expected from ray tracing? Is the sign of the magnification correct? Are the object and image distances reasonable?

Departure from the Small-Angle Approximation

The small-angle approximation is a cornerstone of the above discussion of image formation by a spherical mirror. When this approximation is violated, then the image created by a spherical mirror becomes distorted. Such distortion is called **aberration**. Here we briefly discuss two specific types of aberrations: spherical aberration and coma.

Spherical aberration

Consider a broad beam of parallel rays impinging on a spherical mirror, as shown in [link].



(a) With spherical aberration, the rays that are farther from the optical axis and the rays that are closer to the optical axis are focused at different points. Notice that the aberration gets worse for rays farther from the optical axis. (b) For comatic aberration, parallel rays that are not parallel to the optical axis are focused at different heights and at different focal lengths, so the image contains a “tail” like a comet (which is “coma” in Latin). Note that the colored rays are only to facilitate viewing; the colors do not indicate the color of the light.

The farther from the optical axis the rays strike, the worse the spherical mirror approximates a parabolic mirror. Thus, these rays are not focused at the same point as rays that are near the optical axis, as shown in the figure. Because of **spherical aberration**, the image of an extended object in a spherical mirror will be blurred. Spherical aberrations are characteristic of the mirrors and lenses that we consider in the following section of this chapter (more sophisticated mirrors and lenses are needed to eliminate spherical aberrations).

Coma or comatic aberration

Coma is similar to spherical aberration, but arises when the incoming rays are not parallel to the optical axis, as shown in part (b) of [link]. Recall that the small-angle approximation holds for spherical mirrors that are small compared to their radius. In this case, spherical mirrors are good approximations of parabolic mirrors. Parabolic mirrors focus all rays that are parallel to the optical axis at the focal point. However, parallel rays that are *not* parallel to the optical axis are focused at different heights and at different focal lengths, as show in part (b) of [link]. Because a spherical mirror is symmetric about the optical axis, the various colored rays in this figure create circles of the corresponding color on the focal plane.

Although a spherical mirror is shown in part (b) of [link], comatic aberration occurs also for parabolic mirrors—it does not result from a breakdown in the small-angle approximation. Spherical aberration, however, occurs only for spherical mirrors and is a result of a breakdown in the small-angle approximation. We will discuss both coma and spherical aberration later in this chapter, in connection with telescopes.

Summary

- Spherical mirrors may be concave (converging) or convex (diverging).
- The focal length of a spherical mirror is one-half of its radius of curvature: $f = R/2$.
- The mirror equation and ray tracing allow you to give a complete description of an image formed by a spherical mirror.
- Spherical aberration occurs for spherical mirrors but not parabolic mirrors; comatic aberration occurs for both types of mirrors.

Conceptual Questions

Exercise:

Problem: At what distance is an image always located: at d_o, d_i, or f?

Exercise:

Problem:

Under what circumstances will an image be located at the focal point of a spherical lens or mirror?

Solution:

when the object is at infinity; see the mirror equation

Exercise:

Problem:

What is meant by a negative magnification? What is meant by a magnification whose absolute value is less than one?

Exercise:

Problem:

Can an image be larger than the object even though its magnification is negative? Explain.

Solution:

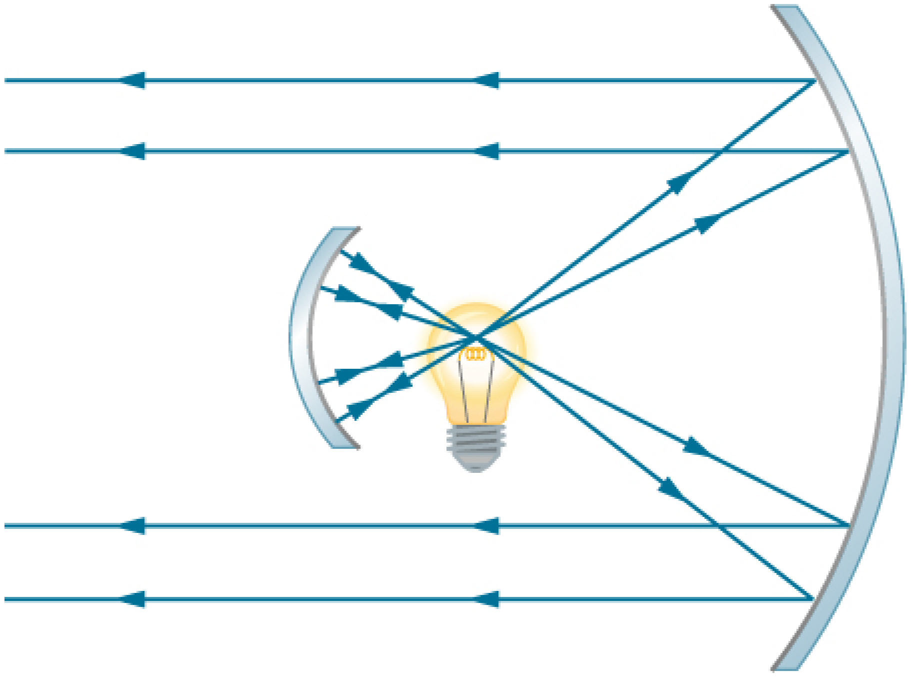
Yes, negative magnification simply means that the image is upside down; this does not prevent the image from being larger than the object. For instance, for a concave mirror, if distance to the object is larger than one focal distance but smaller than two focal distances the image will be inverted and magnified.

Problems

Exercise:

Problem:

The following figure shows a light bulb between two spherical mirrors. One mirror produces a beam of light with parallel rays; the other keeps light from escaping without being put into the beam. Where is the filament of the light in relation to the focal point or radius of curvature of each mirror?



Solution:

It is in the focal point of the big mirror and at the center of curvature of the small mirror.

Exercise:

Problem:

Why are diverging mirrors often used for rearview mirrors in vehicles? What is the main disadvantage of using such a mirror compared with a flat one?

Exercise:

Problem:

Some telephoto cameras use a mirror rather than a lens. What radius of curvature mirror is needed to replace a 800 mm-focal length telephoto lens?

Solution:

$$f=\frac{R}{2}\Rightarrow R=+1.60\,\text{m}$$

Exercise:

Problem:

Calculate the focal length of a mirror formed by the shiny back of a spoon that has a 3.00 cm radius of curvature.

Exercise:

Problem:

Electric room heaters use a concave mirror to reflect infrared (IR) radiation from hot coils. Note that IR radiation follows the same law of reflection as visible light. Given that the mirror has a radius of curvature of 50.0 cm and produces an image of the coils 3.00 m away from the mirror, where are the coils?

Solution:

$d_o = 27.3\,\text{cm}$

Exercise:

Problem:

Find the magnification of the heater element in the previous problem. Note that its large magnitude helps spread out the reflected energy.

Exercise:

Problem:

What is the focal length of a makeup mirror that produces a magnification of 1.50 when a person's face is 12.0 cm away? Explicitly show how you follow the steps in the [link].

Solution:

Step 1: Image formation by a mirror is involved.
Step 2: Draw the problem set up when possible.
Step 3: Use thin-lens equations to solve this problem.
Step 4: Find f.
Step 5: Given: $m = 1.50, d_o = 0.120\,\text{m}$.
Step 6: No ray tracing is needed.
Step 7: Using $m = \frac{d_i}{d_o}$, $d_i = -0.180\,\text{m}$. Then, $f = 0.360\,\text{m}$.
Step 8: The image is virtual because the image distance is negative. The focal length is positive, so the mirror is concave.

Exercise:

Problem:

A shopper standing 3.00 m from a convex security mirror sees his image with a magnification of 0.250. (a) Where is his image? (b) What is the focal length of the mirror? (c) What is its radius of curvature?

Exercise:

Problem:

An object 1.50 cm high is held 3.00 cm from a person’s cornea, and its reflected image is measured to be 0.167 cm high. (a) What is the magnification? (b) Where is the image? (c) Find the radius of curvature of the convex mirror formed by the cornea. (Note that this technique is used by optometrists to measure the curvature of the cornea for contact lens fitting. The instrument used is called a keratometer, or curve measurer.)

Solution:

a. for a convex mirror $d_{\text{i}} < 0 \Rightarrow m > 0$. $m = +0.111$; b.
$d_{\text{i}} = -0.334\,\text{cm}$ (behind the cornea);
c. $f = -0.376\,\text{cm}$, so that $R = -0.752\,\text{cm}$

Exercise:

Problem:

Ray tracing for a flat mirror shows that the image is located a distance behind the mirror equal to the distance of the object from the mirror. This is stated as $d_{\text{i}} = -d_{\text{o}}$, since this is a negative image distance (it is a virtual image). What is the focal length of a flat mirror?

Exercise:

Problem:

Show that, for a flat mirror, $h_{\text{i}} = h_{\text{o}}$, given that the image is the same distance behind the mirror as the distance of the object from the mirror.

Solution:

$m = \frac{h_{\text{i}}}{h_{\text{o}}} = -\frac{d_{\text{i}}}{d_{\text{o}}} = -\frac{-d_{\text{o}}}{d_{\text{o}}} = \frac{d_{\text{o}}}{d_{\text{o}}} = 1 \Rightarrow h_{\text{i}} = h_{\text{o}}$

Exercise:

Problem:

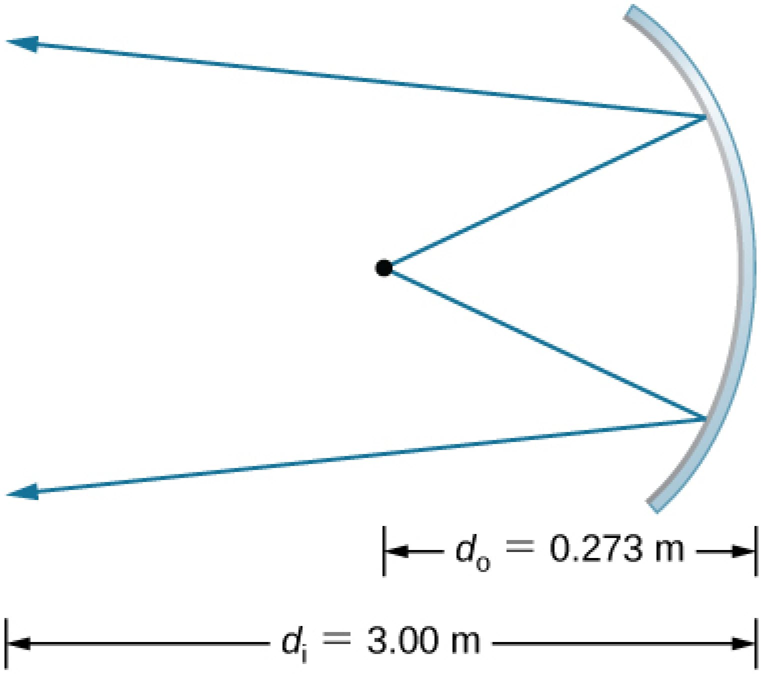
Use the law of reflection to prove that the focal length of a mirror is half its radius of curvature. That is, prove that $f = R/2$. Note this is true for a spherical mirror only if its diameter is small compared with its radius of curvature.

Exercise:

Problem:

Referring to the electric room heater considered in problem 5, calculate the intensity of IR radiation in $\mathrm{W/m^2}$ projected by the concave mirror on a person 3.00 m away. Assume that the heating element radiates 1500 W and has an area of $100\ \mathrm{cm^2}$, and that half of the radiated power is reflected and focused by the mirror.

Solution:



$$m = -11.0$$
$$A' = 0.110\,\mathrm{m}^2$$
$$I = 6.82\,\mathrm{kW/m}^2$$

Exercise:

Problem:

Two mirrors are inclined at an angle of 60° and an object is placed at a point that is equidistant from the two mirrors. Use a protractor to draw rays accurately and locate all images. You may have to draw several figures so that that rays for different images do not clutter your drawing.

Exercise:

Problem:

Two parallel mirrors are facing each other and are separated by a distance of 3 cm. A point object is placed between the mirrors 1 cm from one of the mirrors. Find the coordinates of all the images.

Solution:

$$x_{2m} = -x_{2m-1}, \quad (m = 1, 2, 3, ...),$$
$$x_{2m+1} = b - x_{2m}, \quad (m = 0, 1, 2, ...), \text{ with } x_0 = a.$$

Glossary

aberration
: distortion in an image caused by departures from the small-angle approximation

coma
: similar to spherical aberration, but arises when the incoming rays are not parallel to the optical axis

concave mirror
spherical mirror with its reflecting surface on the inner side of the sphere; the mirror forms a “cave”

convex mirror
spherical mirror with its reflecting surface on the outer side of the sphere

curved mirror
mirror formed by a curved surface, such as spherical, elliptical, or parabolic

focal length
distance along the optical axis from the focal point to the optical element that focuses the light rays

focal point
for a converging lens or mirror, the point at which converging light rays cross; for a diverging lens or mirror, the point from which diverging light rays appear to originate

linear magnification
ratio of image height to object height

optical axis
axis about which the mirror is rotationally symmetric; you can rotate the mirror about this axis without changing anything

small-angle approximation
approximation that is valid when the size of a spherical mirror is significantly smaller than the mirror’s radius; in this approximation, spherical aberration is negligible and the mirror has a well-defined focal point

spherical aberration
distortion in the image formed by a spherical mirror when rays are not all focused at the same point

vertex
: point where the mirror’s surface intersects with the optical axis

Images Formed by Refraction

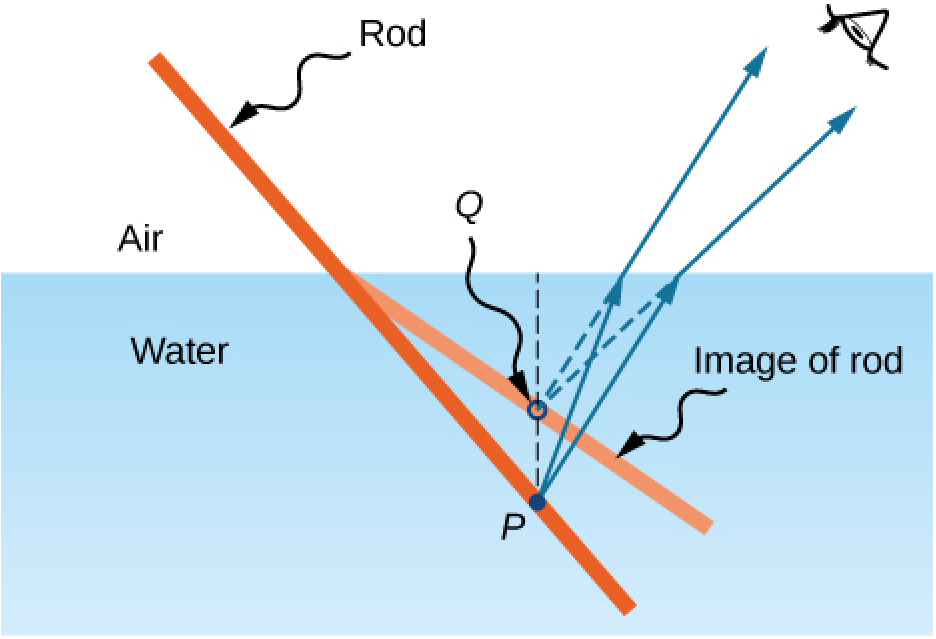
By the end of this section, you will be able to:

- Describe image formation by a single refracting surface
- Determine the location of an image and calculate its properties by using a ray diagram
- Determine the location of an image and calculate its properties by using the equation for a single refracting surface

When rays of light propagate from one medium to another, these rays undergo refraction, which is when light waves are bent at the interface between two media. The refracting surface can form an image in a similar fashion to a reflecting surface, except that the law of refraction (Snell’s law) is at the heart of the process instead of the law of reflection.

Refraction at a Plane Interface—Apparent Depth

If you look at a straight rod partially submerged in water, it appears to bend at the surface ([link]). The reason behind this curious effect is that the image of the rod inside the water forms a little closer to the surface than the actual position of the rod, so it does not line up with the part of the rod that is above the water. The same phenomenon explains why a fish in water appears to be closer to the surface than it actually is.

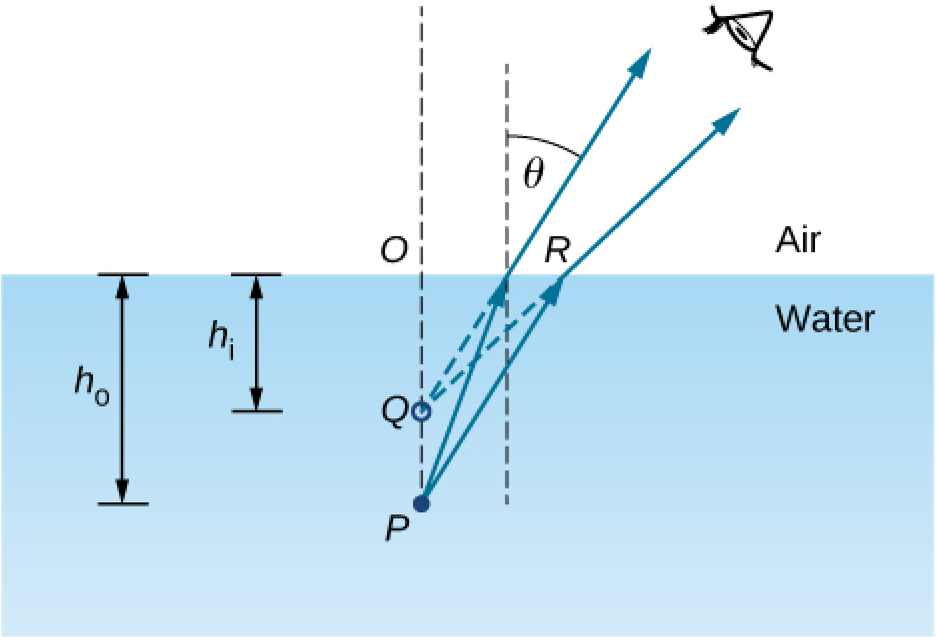


Bending of a rod at a water-air interface. Point *P* on the rod appears to be at point *Q*, which is where the image of point *P* forms due to refraction at the air-water interface.

To study image formation as a result of refraction, consider the following questions:

1. What happens to the rays of light when they enter or pass through a different medium?
2. Do the refracted rays originating from a single point meet at some point or diverge away from each other?

To be concrete, we consider a simple system consisting of two media separated by a plane interface ([link]). The object is in one medium and the observer is in the other. For instance, when you look at a fish from above the water surface, the fish is in medium 1 (the water) with refractive index 1.33, and your eye is in medium 2 (the air) with refractive index 1.00, and the surface of the water is the interface. The depth that you “see” is the image height h_i and is called the **apparent depth**. The actual depth of the fish is the object height h_o.



Apparent depth due to refraction. The real object at point P creates an image at point Q. The image is not at the same depth as the object, so the observer sees the image at an "apparent depth."

The apparent depth h_i depends on the angle at which you view the image. For a view from above (the so-called "normal" view), we can approximate the refraction angle θ to be small, and replace sin θ in Snell's law by tan θ. With this approximation, you can use the triangles ΔOPR and ΔOQR to show that the apparent depth is given by

Note:
Equation:

$$h_\mathrm{i} = \left(\frac{n_2}{n_1}\right)h_\mathrm{o}.$$

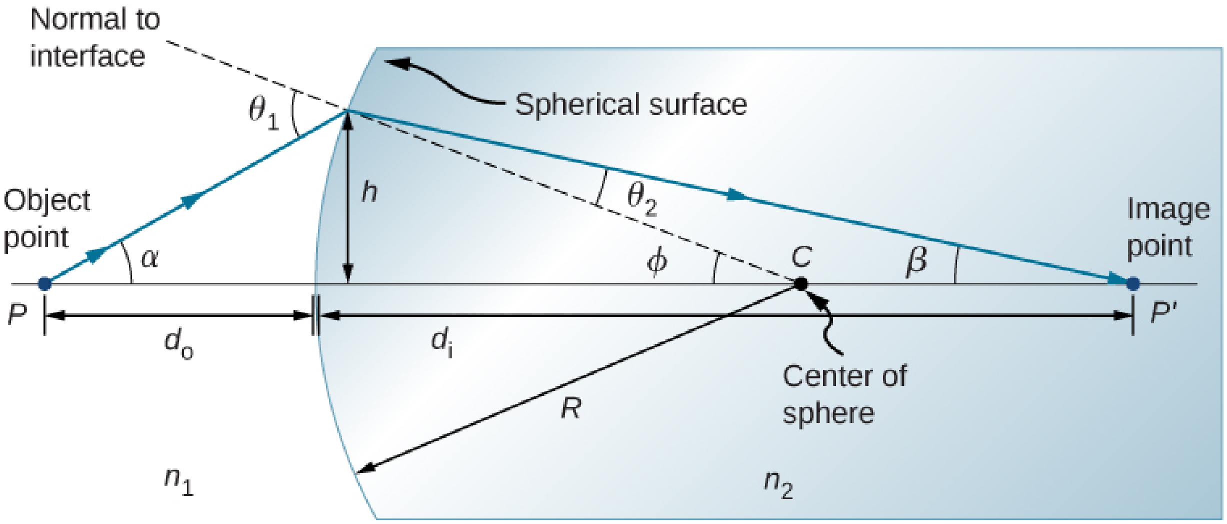
The derivation of this result is left as an exercise. Thus, a fish appears at 3/4 of the real depth when viewed from above.

Refraction at a Spherical Interface

Spherical shapes play an important role in optics primarily because high-quality spherical shapes are far easier to manufacture than other curved surfaces. To study refraction at a single spherical surface, we assume that the medium with the spherical surface at one end continues indefinitely (a “semi-infinite” medium).

Refraction at a convex surface

Consider a point source of light at point *P* in front of a convex surface made of glass (see [link]). Let *R* be the radius of curvature, n_1 be the refractive index of the medium in which object point *P* is located, and n_2 be the refractive index of the medium with the spherical surface. We want to know what happens as a result of refraction at this interface.



Refraction at a convex surface $(n_2 > n_1)$.

Because of the symmetry involved, it is sufficient to examine rays in only one plane. The figure shows a ray of light that starts at the object point P, refracts at the interface, and goes through the image point P'. We derive a formula relating the object distance d_o, the image distance d_i, and the radius of curvature R.

Applying Snell's law to the ray emanating from point P gives $n_1 \sin\theta_1 = n_2 \sin\theta_2$. We work in the small-angle approximation, so $\sin\theta \approx \theta$ and Snell's law then takes the form

Equation:

$$n_1\theta_1 \approx n_2\theta_2.$$

From the geometry of the figure, we see that

Equation:

$$\theta_1 = \alpha + \phi, \quad \theta_2 = \phi - \beta.$$

Inserting these expressions into Snell's law gives

Equation:

$$n_1(\alpha+\phi)\approx n_2(\phi-\beta).$$

Using the diagram, we calculate the tangent of the angles $\alpha, \beta,$ and ϕ:

Equation:

$$\tan\alpha\approx\frac{h}{d_{\text{o}}},\quad \tan\beta\approx\frac{h}{d_{\text{i}}},\quad \tan\phi\approx\frac{h}{R}.$$

Again using the small-angle approximation, we find that $\tan\theta\approx\theta$, so the above relationships become

Equation:

$$\alpha\approx\frac{h}{d_{\text{o}}},\quad \beta\approx\frac{h}{d_{\text{i}}},\quad \phi\approx\frac{h}{R}.$$

Putting these angles into Snell’s law gives

Equation:

$$n_1\left(\frac{h}{d_{\text{o}}}+\frac{h}{R}\right)=n_2\left(\frac{h}{R}-\frac{h}{d_{\text{i}}}\right).$$

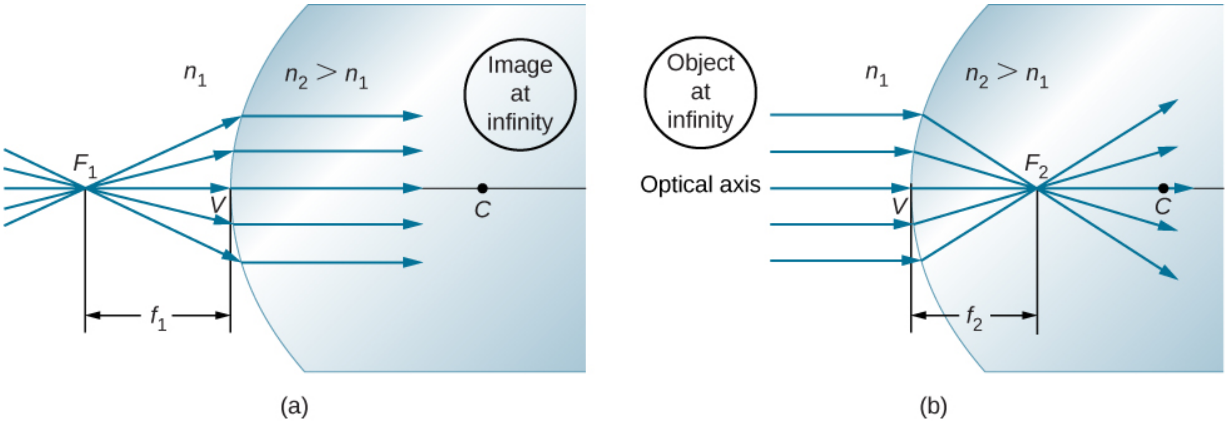
We can write this more conveniently as

Note:

Equation:

$$\frac{n_1}{d_{\text{o}}}+\frac{n_2}{d_{\text{i}}}=\frac{n_2-n_1}{R}.$$

If the object is placed at a special point called the **first focus**, or the **object focus** F_1, then the image is formed at infinity, as shown in part (a) of [link].



(a) First focus (called the “object focus”) for refraction at a convex surface. (b) Second focus (called “image focus”) for refraction at a convex surface.

We can find the location f_1 of the first focus F_1 by setting $d_\text{i} = \infty$ in the preceding equation.

Equation:

$$\frac{n_1}{f_1} + \frac{n_2}{\infty} = \frac{n_2 - n_1}{R}$$

Equation:

$$f_1 = \frac{n_1 R}{n_2 - n_1}$$

Similarly, we can define a **second focus** or **image focus** F_2 where the image is formed for an object that is far away [part (b)]. The location of the second focus F_2 is obtained from [link] by setting $d_\text{o} = \infty$:

Equation:

$$\frac{n_1}{\infty} + \frac{n_2}{f_2} = \frac{n_2 - n_1}{R}$$

Equation:

$$f_2 = \frac{n_2 R}{n_2 - n_1}.$$

Note that the object focus is at a different distance from the vertex than the image focus because $n_1 \neq n_2$.

Sign convention for single refracting surfaces

Although we derived this equation for refraction at a convex surface, the same expression holds for a concave surface, provided we use the following sign convention:

1. $R > 0$ if surface is convex toward object; otherwise, $R < 0$.
2. $d_\mathrm{i} > 0$ if image is real and on opposite side from the object; otherwise, $d_\mathrm{i} < 0$.

Summary

This section explains how a single refracting interface forms images.

- When an object is observed through a plane interface between two media, then it appears at an apparent distance h_i that differs from the actual distance h_o: $h_\mathrm{i} = (n_2/n_1)h_\mathrm{o}$.
- An image is formed by the refraction of light at a spherical interface between two media of indices of refraction n_1 and n_2.
- Image distance depends on the radius of curvature of the interface, location of the object, and the indices of refraction of the media.

Conceptual Questions

Exercise:

Problem:

Derive the formula for the apparent depth of a fish in a fish tank using Snell’s law.

Exercise:

Problem:

Use a ruler and a protractor to find the image by refraction in the following cases. Assume an air-glass interface. Use a refractive index of 1 for air and of 1.5 for glass. (*Hint*: Use Snell’s law at the interface.)

(a) A point object located on the axis of a concave interface located at a point within the focal length from the vertex.

(b) A point object located on the axis of a concave interface located at a point farther than the focal length from the vertex.

(c) A point object located on the axis of a convex interface located at a point within the focal length from the vertex.

(d) A point object located on the axis of a convex interface located at a point farther than the focal length from the vertex.

(e) Repeat (a)–(d) for a point object off the axis.

Solution:

answers may vary

Problems

Exercise:

Problem:

An object is located in air 30 cm from the vertex of a concave surface made of glass with a radius of curvature 10 cm. Where does the image by refraction form and what is its magnification? Use $n_{\text{air}} = 1$ and $n_{\text{glass}} = 1.5$.

Exercise:

Problem:

An object is located in air 30 cm from the vertex of a convex surface made of glass with a radius of curvature 80 cm. Where does the image by refraction form and what is its magnification?

Solution:

$d_{\text{i}} = -55 \text{ cm}; m = +1.8$

Exercise:

Problem:

An object is located in water 15 cm from the vertex of a concave surface made of glass with a radius of curvature 10 cm. Where does the image by refraction form and what is its magnification? Use $n_{\text{water}} = 4/3$ and $n_{\text{glass}} = 1.5$.

Exercise:

Problem:

An object is located in water 30 cm from the vertex of a convex surface made of Plexiglas with a radius of curvature of 80 cm. Where does the image form by refraction and what is its magnification? $n_{\text{water}} = 4/3$ and $n_{\text{Plexiglas}} = 1.65$.

Solution:

$d_{\text{i}} = -41 \text{ cm}, m = 1.4$

Exercise:

Problem:

An object is located in air 5 cm from the vertex of a concave surface made of glass with a radius of curvature 20 cm. Where does the image form by refraction and what is its magnification? Use $n_{\text{air}} = 1$ and $n_{\text{glass}} = 1.5$.

Exercise:

Problem:

Derive the spherical interface equation for refraction at a concave surface. (*Hint*: Follow the derivation in the text for the convex surface.)

Solution:

proof

Glossary

apparent depth
: depth at which an object is perceived to be located with respect to an interface between two media

first focus or object focus
: object located at this point will result in an image created at infinity on the opposite side of a spherical interface between two media

second focus or image focus
: for a converging interface, the point where a bundle of parallel rays refracting at a spherical interface; for a diverging interface, the point at which the backward continuation of the refracted rays will converge between two media will focus

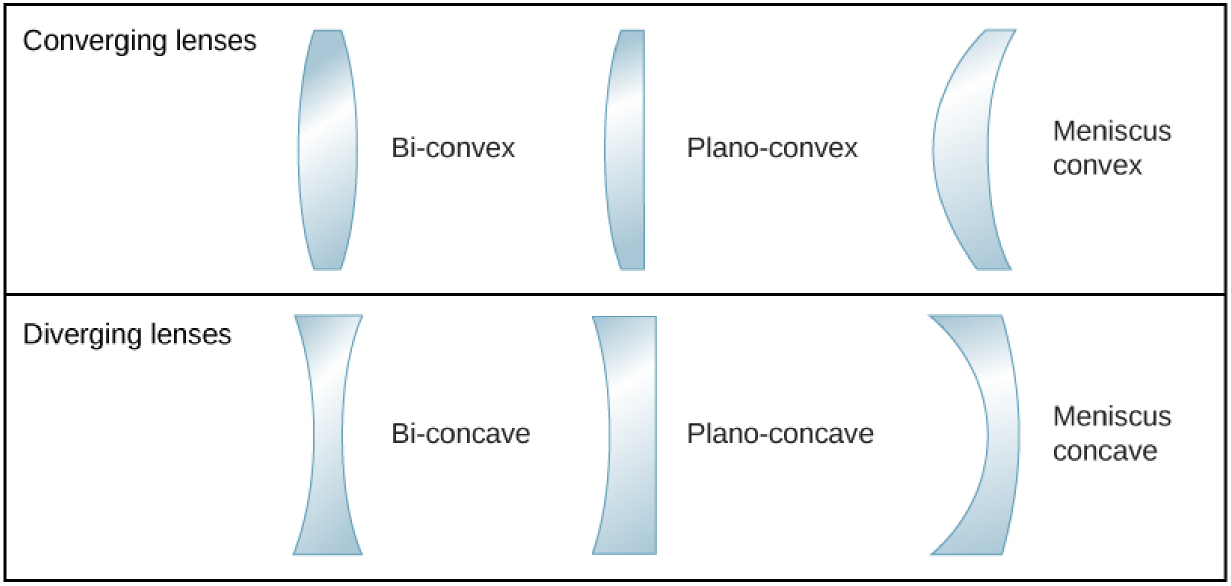
Thin Lenses

By the end of this section, you will be able to:

- Use ray diagrams to locate and describe the image formed by a lens
- Employ the thin-lens equation to describe and locate the image formed by a lens

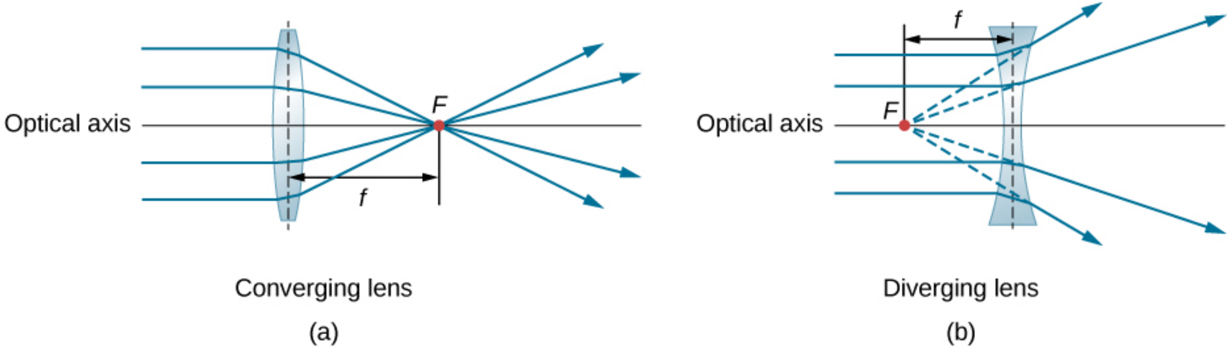
Lenses are found in a huge array of optical instruments, ranging from a simple magnifying glass to a camera's zoom lens to the eye itself. In this section, we use the Snell's law to explore the properties of lenses and how they form images.

The word "lens" derives from the Latin word for a lentil bean, the shape of which is similar to a convex lens. However, not all lenses have the same shape. [link] shows a variety of different lens shapes. The vocabulary used to describe lenses is the same as that used for spherical mirrors: The axis of symmetry of a lens is called the optical axis, where this axis intersects the lens surface is called the vertex of the lens, and so forth.



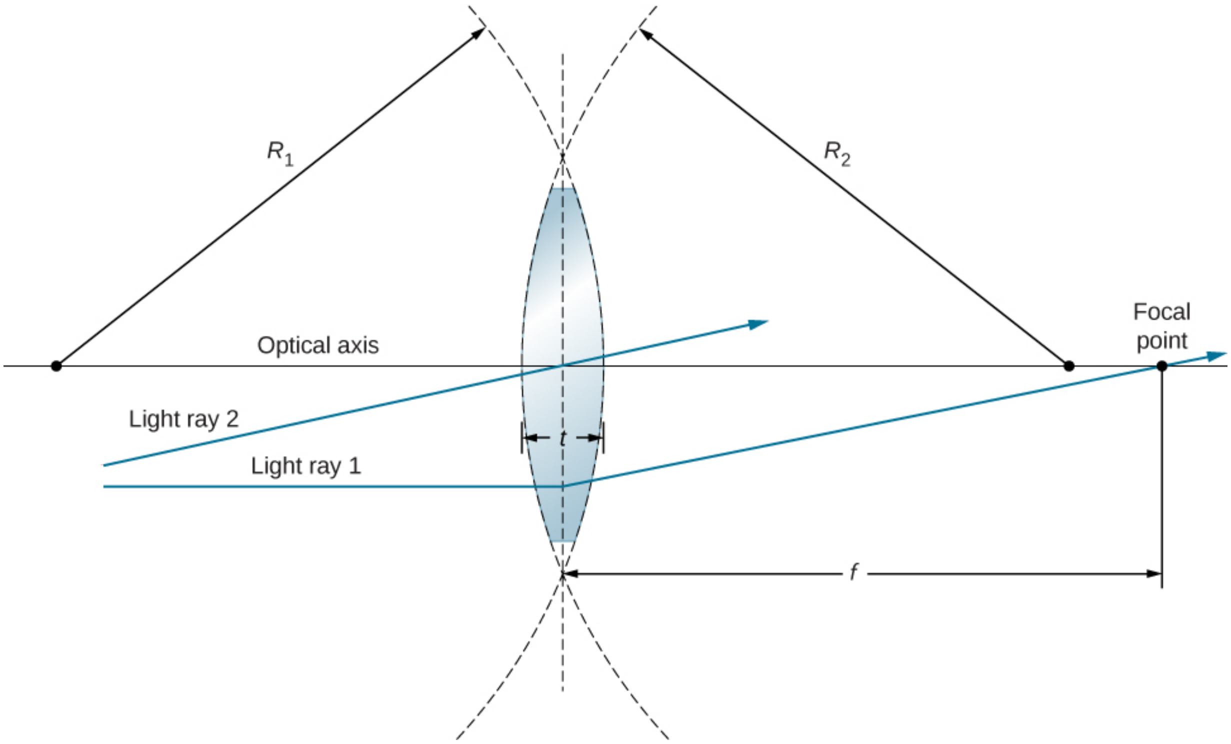
Various types of lenses: Note that a converging lens has a thicker "waist," whereas a diverging lens has a thinner waist.

A **convex** or **converging lens** is shaped so that all light rays that enter it parallel to its optical axis intersect (or focus) at a single point on the optical axis on the opposite side of the lens, as shown in part (a) of [link]. Likewise, a **concave** or **diverging lens** is shaped so that all rays that enter it parallel to its optical axis diverge, as shown in part (b). To understand more precisely how a lens manipulates light, look closely at the top ray that goes through the converging lens in part (a). Because the index of refraction of the lens is greater than that of air, Snell's law tells us that the ray is bent toward the perpendicular to the interface as it enters the lens. Likewise, when the ray exits the lens, it is bent away from the perpendicular. The same reasoning applies to the diverging lenses, as shown in part (b). The overall effect is that light rays are bent toward the optical axis for a converging lens and away from the optical axis for diverging lenses. For a converging lens, the point at which the rays cross is the focal point *F* of the lens. For a diverging lens, the point from which the rays appear to originate is the (virtual) focal point. The distance from the center of the lens to its focal point is the focal length *f* of the lens.



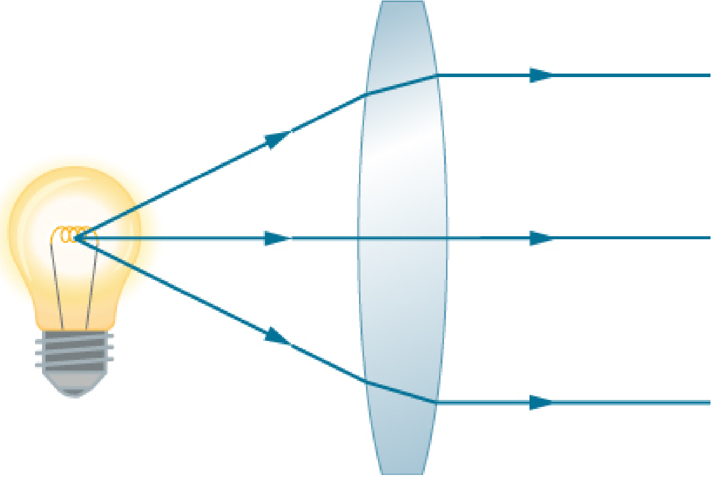
Rays of light entering (a) a converging lens and (b) a diverging lens, parallel to its axis, converge at its focal point *F*. The distance from the center of the lens to the focal point is the lens's focal length *f*. Note that the light rays are bent upon entering and exiting the lens, with the overall effect being to bend the rays toward the optical axis.

A lens is considered to be thin if its thickness *t* is much less than the radii of curvature of both surfaces, as shown in [link]. In this case, the rays may be considered to bend once at the center of the lens. For the case drawn in the figure, light ray 1 is parallel to the optical axis, so the outgoing ray is bent once at the center of the lens and goes through the focal point. Another important characteristic of thin lenses is that light rays that pass through the center of the lens are undeviated, as shown by light ray 2.



In the thin-lens approximation, the thickness *t* of the lens is much, much less than the radii R_1 and R_2 of curvature of the surfaces of the lens. Light rays are considered to bend at the center of the lens, such as light ray 1. Light ray 2 passes through the center of the lens and is undeviated in the thin-lens approximation.

As noted in the initial discussion of Snell’s law, the paths of light rays are exactly reversible. This means that the direction of the arrows could be reversed for all of the rays in [link]. For example, if a point-light source is placed at the focal point of a convex lens, as shown in [link], parallel light rays emerge from the other side.



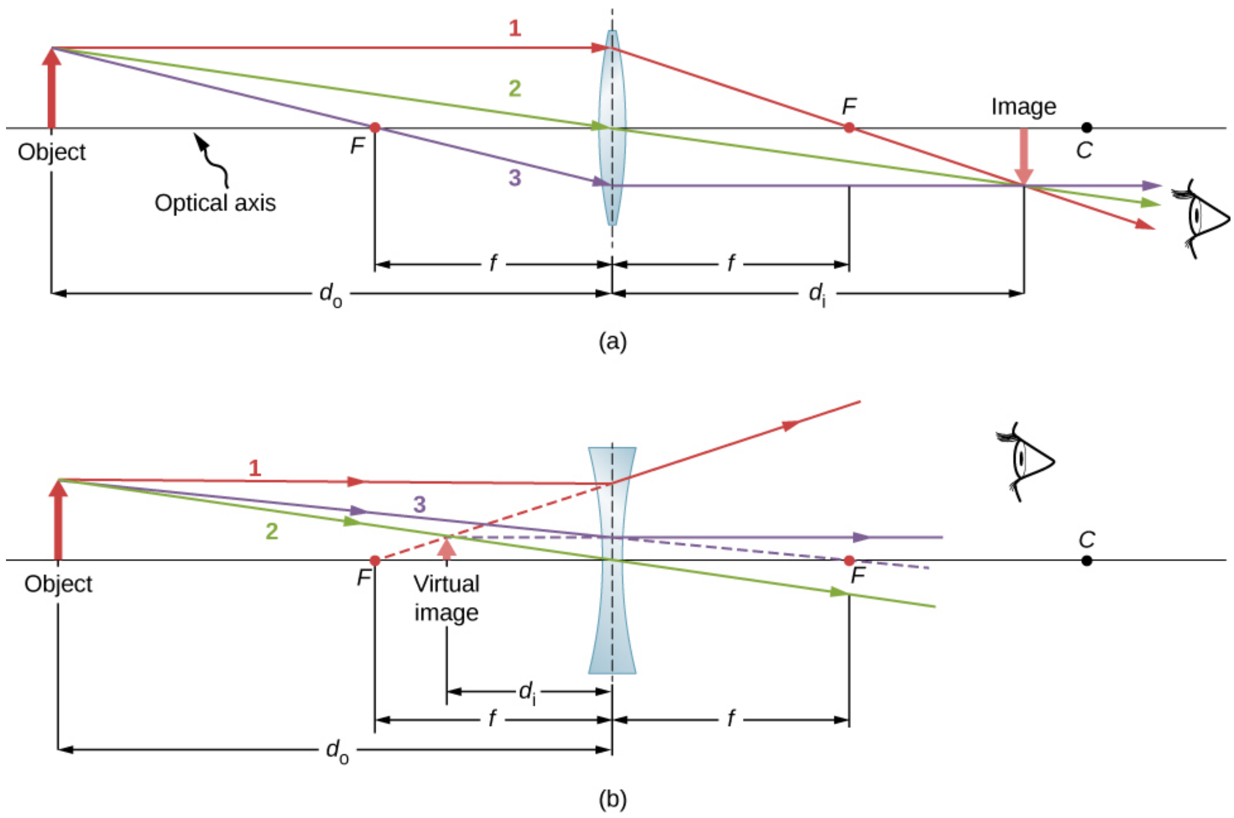
A small light source, like a light bulb filament, placed at the focal point of a convex lens results in parallel rays of light emerging from the other side. The paths are exactly the reverse of those shown in [link] in converging and diverging lenses. This technique is used in lighthouses and sometimes in traffic lights to produce a directional beam of light from a source that emits light in all directions.

Ray Tracing and Thin Lenses

Ray tracing is the technique of determining or following (tracing) the paths taken by light rays.

Ray tracing for thin lenses is very similar to the technique we used with spherical mirrors. As for mirrors, ray tracing can accurately describe the operation of a lens. The rules for ray tracing for thin lenses are similar to those of spherical mirrors:

1. A ray entering a converging lens parallel to the optical axis passes through the focal point on the other side of the lens (ray 1 in part (a) of [link]). A ray entering a diverging lens parallel to the optical axis exits along the line that passes through the focal point on the *same* side of the lens (ray 1 in part (b) of the figure).
2. A ray passing through the center of either a converging or a diverging lens is not deviated (ray 2 in parts (a) and (b)).
3. For a converging lens, a ray that passes through the focal point exits the lens parallel to the optical axis (ray 3 in part (a)). For a diverging lens, a ray that approaches along the line that passes through the focal point on the opposite side exits the lens parallel to the axis (ray 3 in part (b)).



Thin lenses have the same focal lengths on either side. (a) Parallel light rays from the object toward a converging lens cross at its focal point on the right. (b) Parallel light rays from the object entering a diverging lens from the left seem to come from the focal point on the left.

Thin lenses work quite well for monochromatic light (i.e., light of a single wavelength). However, for light that contains several wavelengths (e.g., white light), the lenses work less well. The problem is that, as we learned in the previous chapter, the index of refraction of a material depends on the wavelength of light. This phenomenon is responsible for many colorful effects, such as rainbows. Unfortunately, this phenomenon also leads to aberrations in images formed by lenses. In particular, because the focal distance of the lens depends on the index of refraction, it also depends on the wavelength of the incident light. This means that light of different wavelengths will focus at different points, resulting is so-called “chromatic aberrations.” In particular, the edges of an image of a white object will become colored and blurred. Special lenses called doublets are capable of correcting chromatic aberrations. A doublet is formed by gluing together a converging lens and a diverging lens. The combined doublet lens produces significantly reduced chromatic aberrations.

Image Formation by Thin Lenses

We use ray tracing to investigate different types of images that can be created by a lens. In some circumstances, a lens forms a real image, such as when a movie projector casts an image onto a screen. In other cases, the image is a virtual image, which cannot be projected onto a screen. Where, for example, is the image formed by eyeglasses? We use ray tracing for thin lenses to illustrate how they form images, and then we develop equations to analyze quantitatively the properties of thin lenses.

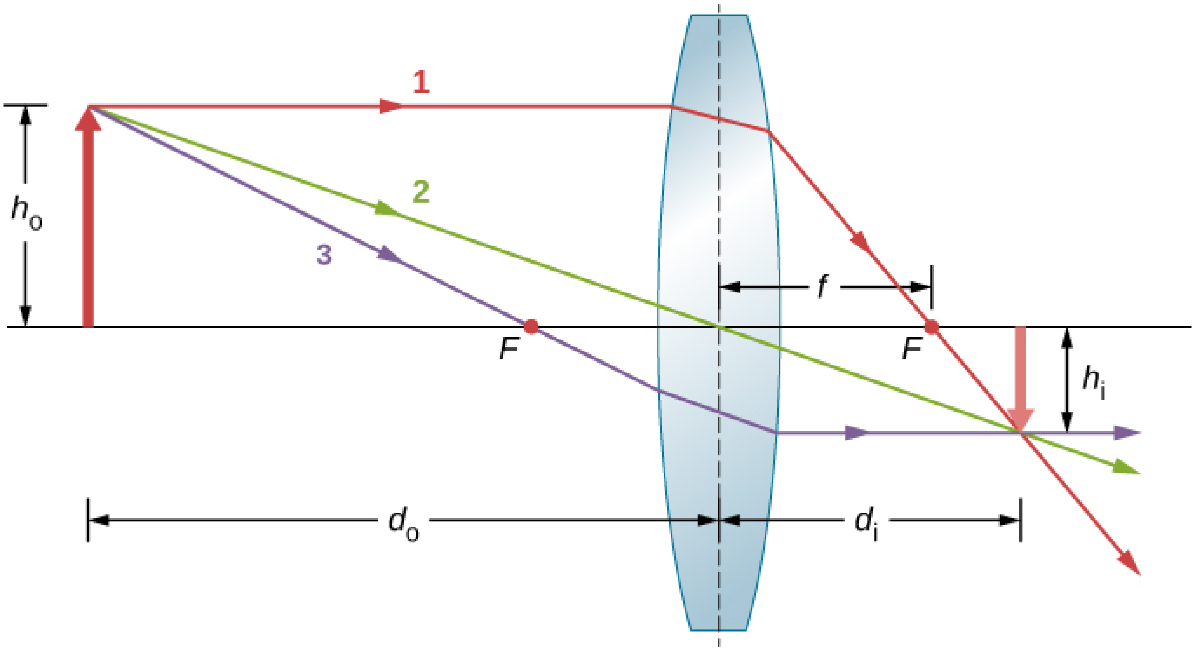
Consider an object some distance away from a converging lens, as shown in [link]. To find the location and size of the image, we trace the paths of selected light rays originating from one point on the object, in this case, the tip of the arrow. The figure shows three rays from many rays that emanate from the tip of the arrow. These three rays can be traced by using the ray-tracing rules given above.

- Ray 1 enters the lens parallel to the optical axis and passes through the focal point on the opposite side (rule 1).
- Ray 2 passes through the center of the lens and is not deviated (rule 2).
- Ray 3 passes through the focal point on its way to the lens and exits the lens parallel to the optical axis (rule 3).

The three rays cross at a single point on the opposite side of the lens. Thus, the image of the tip of the arrow is located at this point. All rays that come from the tip of the arrow and enter the lens are refracted and cross at the point shown.

After locating the image of the tip of the arrow, we need another point of the image to orient the entire image of the arrow. We chose to locate the image base of the arrow, which is on the optical axis. As explained in the section on spherical mirrors, the base will be on the optical axis just above the image of the tip of the arrow (due to the top-bottom symmetry of the lens). Thus, the image spans the optical axis to the (negative) height shown. Rays from another point on the arrow, such as the middle of the arrow, cross at another common point, thus filling in the rest of the image.

Although three rays are traced in this figure, only two are necessary to locate a point of the image. It is best to trace rays for which there are simple ray-tracing rules.

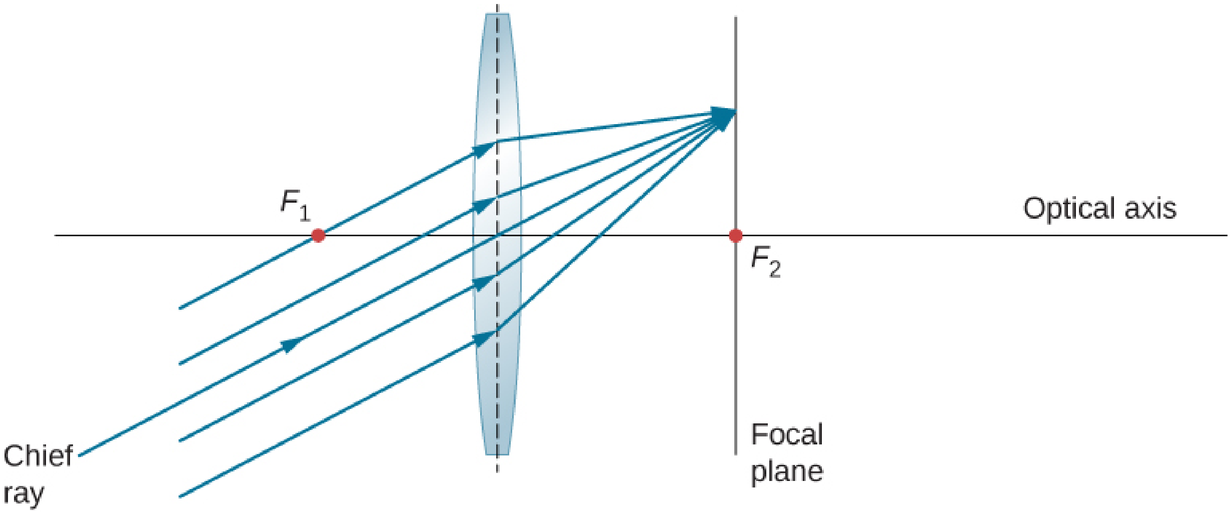


Ray tracing is used to locate the image formed by a lens. Rays originating from the same point on the object are traced—the three chosen rays each follow one of the rules for ray tracing, so that their paths are easy to determine. The image is located at the point where the rays cross. In this case, a real image—one that can be projected on a screen—is formed.

Several important distances appear in the figure. As for a mirror, we define d_o to be the object distance, or the distance of an object from the center of a lens. The image distance d_i is defined to be the distance of the image from the center of a lens. The height of the object and the height of the image are indicated by h_o and h_i, respectively. Images that appear upright relative to the object have positive heights, and those that are inverted have negative heights. By using the rules of ray tracing and making a scale drawing with paper and pencil, like that in [link], we can accurately describe the location and size of an image. But the real benefit of ray tracing is in visualizing how images are formed in a variety of situations.

Oblique Parallel Rays and Focal Plane

We have seen that rays parallel to the optical axis are directed to the focal point of a converging lens. In the case of a diverging lens, they come out in a direction such that they appear to be coming from the focal point on the opposite side of the lens (i.e., the side from which parallel rays enter the lens). What happens to parallel rays that are not parallel to the optical axis ([link])? In the case of a converging lens, these rays do not converge at the focal point. Instead, they come together on another point in the plane called the **focal plane**. The focal plane contains the focal point and is perpendicular to the optical axis. As shown in the figure, parallel rays focus where the ray through the center of the lens crosses the focal plane.



Parallel oblique rays focus on a point in a focal plane.

Thin-Lens Equation

Ray tracing allows us to get a qualitative picture of image formation. To obtain numeric information, we derive a pair of equations from a geometric analysis of ray tracing for thin lenses. These equations, called the thin-lens equation and the lens maker's equation, allow us to quantitatively analyze thin lenses.

Consider the thick bi-convex lens shown in [link]. The index of refraction of the surrounding medium is n_1 (if the lens is in air, then $n_1 = 1.00$) and that of the lens is n_2. The radii of curvatures of the two sides are R_1 and R_2. We wish to find a relation between the object distance d_o, the image distance d_i, and the parameters of the lens.

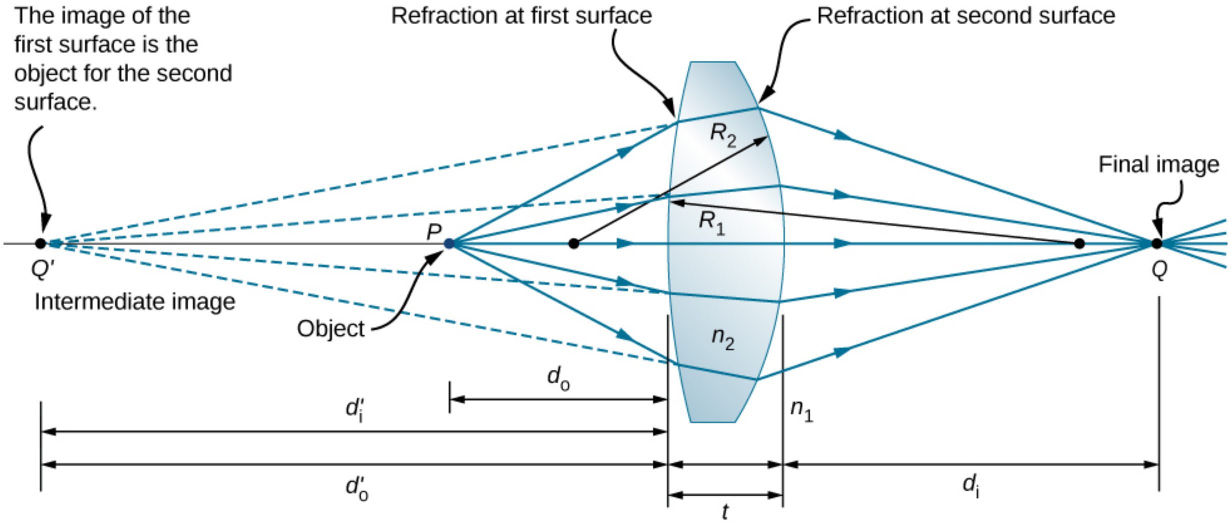

The image of the first surface is the object for the second surface.
Refraction at first surface
Refraction at second surface
Final image
R_2
R_1
P
Q'
Q
Intermediate image
Object
n_2
n_1
d_o
d_i'
d_o'
d_i
t

Figure for deriving the lens maker’s equation. Here, t is the thickness of lens, n_1 is the index of refraction of the exterior medium, and n_2 is the index of refraction of the lens. We take the limit of $t \to 0$ to obtain the formula for a thin lens.

To derive the thin-lens equation, we consider the image formed by the first refracting surface (i.e., left surface) and then use this image as the object for the second refracting surface. In the figure, the image from the first refracting surface is Q', which is formed by extending backwards the rays from inside the lens (these rays result from refraction at the first surface). This is shown by the dashed lines in the figure. Notice that this image is virtual because no rays actually pass through the point Q'. To find the image distance d_{i}' corresponding to the image Q', we use [link]. In this case, the object distance is d_{o}, the image distance is d_{i}', and the radius of curvature is R_1. Inserting these into [link] gives

Equation:

$$\frac{n_1}{d_{\text{o}}} + \frac{n_2}{d_{\text{i}}'} = \frac{n_2 - n_1}{R_1}.$$

The image is virtual and on the same side as the object, so $d_{\text{i}}' < 0$ and $d_{\text{o}} > 0$. The first surface is convex toward the object, so $R_1 > 0$.

To find the object distance for the object Q formed by refraction from the second interface, note that the role of the indices of refraction n_1 and n_2 are interchanged in [link]. In [link], the rays originate in the medium with index n_2, whereas in [link], the rays originate in the medium with index n_1. Thus, we must interchange n_1 and n_2 in [link]. In addition, by consulting again [link], we see that the object distance is d_{o}' and the image distance is d_{i}. The radius of curvature is R_2 Inserting these quantities into [link] gives

Equation:

$$\frac{n_2}{d'_{\mathrm{o}}}+\frac{n_1}{d_{\mathrm{i}}}=\frac{n_1-n_2}{R_2}.$$

The image is real and on the opposite side from the object, so $d_{\mathrm{i}}>0$ and $d'_{\mathrm{o}}>0$. The second surface is convex away from the object, so $R_2<0$. [link] can be simplified by noting that $d'_{\mathrm{o}}=|d'_{\mathrm{i}}|+t$, where we have taken the absolute value because d'_{i} is a negative number, whereas both d'_{o} and *t* are positive. We can dispense with the absolute value if we negate d'_{i}, which gives $d'_{\mathrm{o}}=-d'_{\mathrm{i}}+t$. Inserting this into [link] gives

Equation:

$$\frac{n_2}{-d'_{\mathrm{i}}+t}+\frac{n_1}{d_{\mathrm{i}}}=\frac{n_1-n_2}{R_2}.$$

Summing [link] and [link] gives

Equation:

$$\frac{n_1}{d_{\mathrm{o}}}+\frac{n_1}{d_{\mathrm{i}}}+\frac{n_2}{d'_{\mathrm{i}}}+\frac{n_2}{-d'_{\mathrm{i}}+t}=(n_2-n_1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right).$$

In the **thin-lens approximation**, we assume that the lens is very thin compared to the first image distance, or $t\ll d'_{\mathrm{i}}$ (or, equivalently, $t\ll R_1$ and R_2). In this case, the third and fourth terms on the left-hand side of [link] cancel, leaving us with

Equation:

$$\frac{n_1}{d_{\mathrm{o}}}+\frac{n_1}{d_{\mathrm{i}}}=(n_2-n_1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right).$$

Dividing by n_1 gives us finally

Equation:

$$\frac{1}{d_{\mathrm{o}}}+\frac{1}{d_{\mathrm{i}}}=\left(\frac{n_2}{n_1}-1\right)\left(\frac{1}{R_1}-\frac{1}{R_2}\right).$$

The left-hand side looks suspiciously like the mirror equation that we derived above for spherical mirrors. As done for spherical mirrors, we can use ray tracing and geometry to show that, for a thin lens,

Note:
Equation:

$$\frac{1}{d_{\mathrm{o}}}+\frac{1}{d_{\mathrm{i}}}=\frac{1}{f}$$

where *f* is the focal length of the thin lens (this derivation is left as an exercise). This is the thin-lens equation. The focal length of a thin lens is the same to the left and to the right of the lens. Combining [link] and [link] gives

Note:
Equation:

$$\frac{1}{f}=\left(\frac{n_2}{n_1}-1\right)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

which is called the lens maker's equation. It shows that the focal length of a thin lens depends only of the radii of curvature and the index of refraction of the lens and that of the surrounding medium. For a lens in air, $n_1 = 1.0$ and $n_2 \equiv n$, so the lens maker's equation reduces to
Equation:

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right).$$

Sign conventions for lenses

To properly use the thin-lens equation, the following sign conventions must be obeyed:

1. d_i is positive if the image is on the side opposite the object (i.e., real image); otherwise, d_i is negative (i.e., virtual image).
2. *f* is positive for a converging lens and negative for a diverging lens.
3. *R* is positive for a surface convex toward the object, and negative for a surface concave toward object.

Magnification

By using a finite-size object on the optical axis and ray tracing, you can show that the magnification *m* of an image is

Note:
Equation:

$$m \equiv \frac{h_\text{i}}{h_\text{o}} = -\frac{d_\text{i}}{d_\text{o}}$$

(where the three lines mean “is defined as”). This is exactly the same equation as we obtained for mirrors (see [link]). If $m > 0$, then the image has the same vertical orientation as the object (called an “upright” image). If $m < 0$, then the image has the opposite vertical orientation as the object (called an “inverted” image).

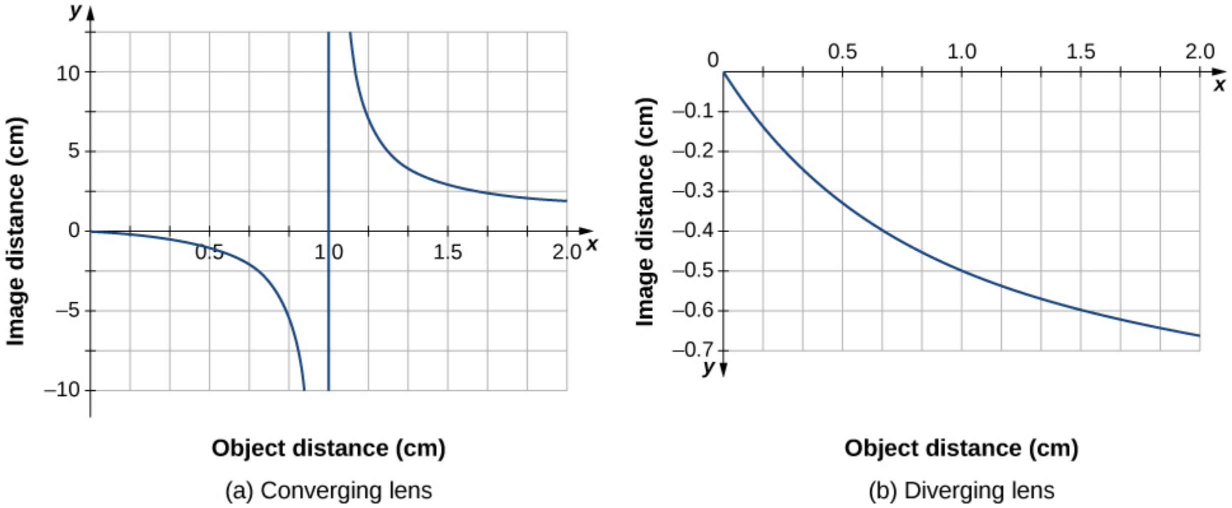
Using the Thin-Lens Equation

The thin-lens equation and the lens maker’s equation are broadly applicable to situations involving thin lenses. We explore many features of image formation in the following examples.

Consider a thin converging lens. Where does the image form and what type of image is formed as the object approaches the lens from infinity? This may be seen by using the thin-lens equation for a given focal length to plot the image distance as a function of object distance. In other words, we plot
Equation:

$$d_{\mathrm{i}} = \left(\frac{1}{f} - \frac{1}{d_{\mathrm{o}}}\right)^{-1}$$

for a given value of f. For $f = 1\ \mathrm{cm}$, the result is shown in part (a) of [link].



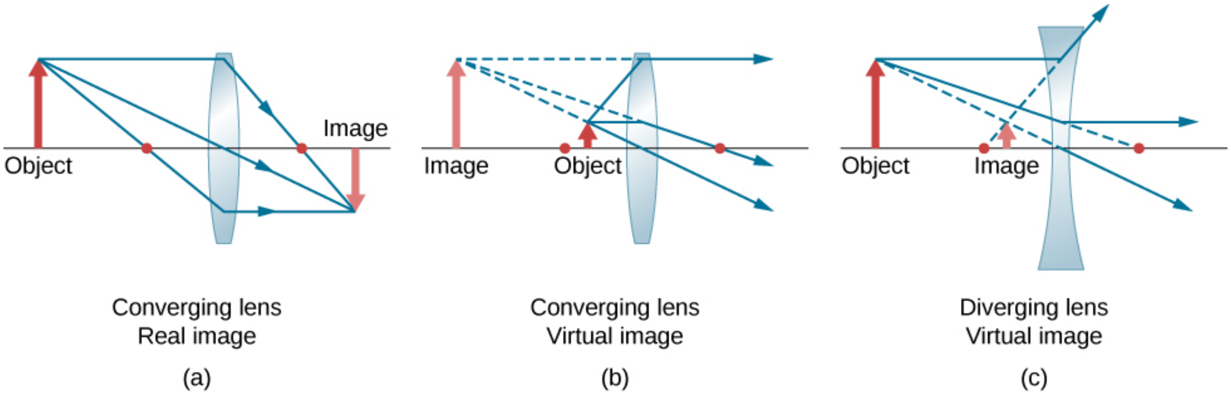
(a) Converging lens

(b) Diverging lens

(a) Image distance for a thin converging lens with $f = 1.0\ \mathrm{cm}$ as a function of object distance. (b) Same thing but for a diverging lens with $f = -1.0\ \mathrm{cm}$.

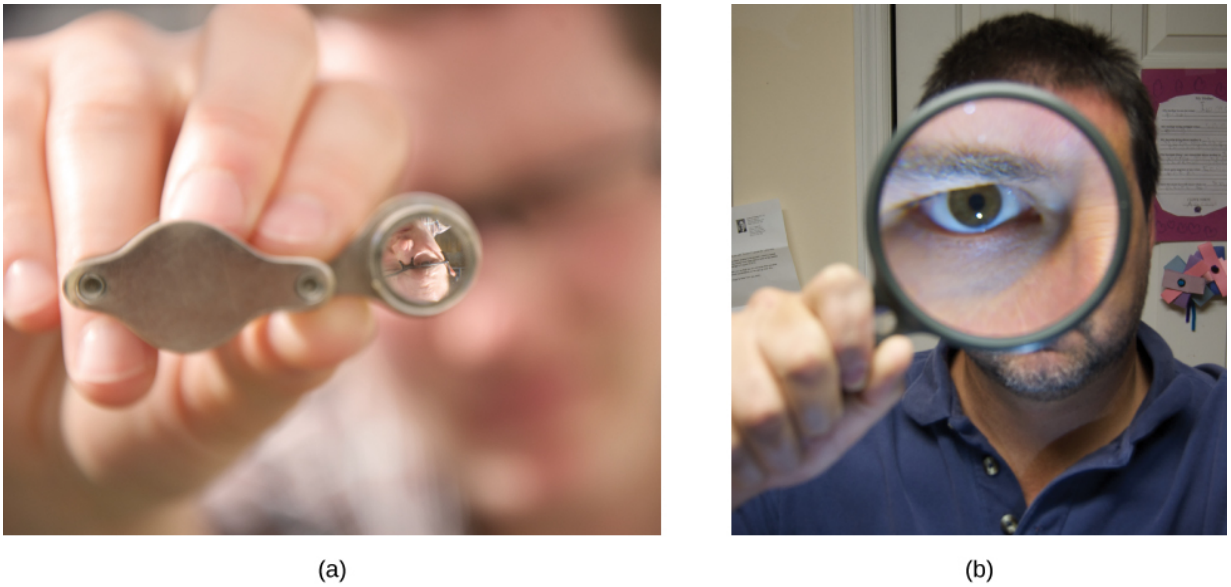
An object much farther than the focal length f from the lens should produce an image near the focal plane, because the second term on the right-hand side of the equation above becomes negligible compared to the first term, so we have $d_{\mathrm{i}} \approx f$. This can be seen in the plot of part (a) of the figure, which shows that the image distance approaches asymptotically the focal length of 1 cm for larger object distances. As the object approaches the focal plane, the image distance diverges to positive infinity. This is expected because an object at the focal plane produces parallel rays that form an image at infinity (i.e., very far from the lens). When the object is farther than the focal length from the lens, the image distance is positive, so the image is real, on the opposite side of the lens from the object, and inverted (because $m = -d_{\mathrm{i}}/d_{\mathrm{o}}$). When the object is closer than the focal length from the lens, the image distance becomes negative, which means that the image is virtual, on the same side of the lens as the object, and upright.

For a thin diverging lens of focal length $f = -1.0\ \mathrm{cm}$, a similar plot of image distance vs. object distance is shown in part (b). In this case, the image distance is negative for all positive object distances, which means that the image is virtual, on the same side of the lens as the object, and upright. These characteristics may also be seen by ray-tracing diagrams (see [link]).



The red dots show the focal points of the lenses. (a) A real, inverted image formed from an object that is farther than the focal length from a converging lens. (b) A virtual, upright image formed from an object that is closer than a focal length from the lens. (c) A virtual, upright image formed from an object that is farther than a focal length from a diverging lens.

To see a concrete example of upright and inverted images, look at [link], which shows images formed by converging lenses when the object (the person’s face in this case) is place at different distances from the lens. In part (a) of the figure, the person’s face is farther than one focal length from the lens, so the image is inverted. In part (b), the person’s face is closer than one focal length from the lens, so the image is upright.


(a)
(b)

(a) When a converging lens is held farther than one focal length from the man’s face, an inverted image is formed. Note that the image is in focus but the face is not, because the image is much closer to the camera taking this photograph than the face. (b) An upright image of the man’s face is produced when a converging lens is held at less than one focal length from his face. (credit a: modification of work by “DaMongMan”/Flickr; credit b: modification of work by Casey Fleser)

Work through the following examples to better understand how thin lenses work.

Note:
Problem-Solving Strategy: Lenses
Step 1. Determine whether ray tracing, the thin-lens equation, or both would be useful. Even if ray tracing is not used, a careful sketch is always very useful. Write symbols and values on the sketch.
Step 2. Identify what needs to be determined in the problem (identify the unknowns).
Step 3. Make a list of what is given or can be inferred from the problem (identify the knowns).
Step 4. If ray tracing is required, use the ray-tracing rules listed near the beginning of this section.
Step 5. Most quantitative problems require the use of the thin-lens equation and/or the lens maker’s equation. Solve these for the unknowns and insert the given quantities or use both together to find two unknowns.

Step 7. Check to see if the answer is reasonable. Are the signs correct? Is the sketch or ray tracing consistent with the calculation?

Example:

Using the Lens Maker’s Equation

Find the radius of curvature of a biconcave lens symmetrically ground from a glass with index of refractive 1.55 so that its focal length in air is 20 cm (for a biconcave lens, both surfaces have the same radius of curvature).

Strategy

Use the thin-lens form of the lens maker’s equation:

Equation:

$$\frac{1}{f}=\left(\frac{n_2}{n_1}-1\right)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

where $R_1 < 0$ and $R_2 > 0$. Since we are making a symmetric biconcave lens, we have $|R_1| = |R_2|$.

Solution

We can determine the radius *R* of curvature from

Equation:

$$\frac{1}{f}=\left(\frac{n_2}{n_1}-1\right)\left(-\frac{2}{R}\right).$$

Solving for *R* and inserting $f=-20\,\text{cm}, n_2=1.55, \text{ and } n_1=1.00$ gives

Equation:

$$R=-2f\left(\frac{n_2}{n_1}-1\right)=-2(-20\,\text{cm})\left(\frac{1.55}{1.00}-1\right)=22\,\text{cm}.$$

Example:

Converging Lens and Different Object Distances

Find the location, orientation, and magnification of the image for an 3.0 cm high object at each of the following positions in front of a convex lens of focal length 10.0 cm. (a) $d_\text{o}=50.0\,\text{cm}$, (b) $d_\text{o}=5.00\,\text{cm}$, and (c) $d_\text{o}=20.0\,\text{cm}$.

Strategy

We start with the thin-lens equation $\frac{1}{d_\text{i}}+\frac{1}{d_\text{o}}=\frac{1}{f}$. Solve this for the image distance d_i and insert the given object distance and focal length.

Solution

a. For $d_\text{o} = 50\text{ cm}$, $f = +10\text{ cm}$, this gives
Equation:

$$\begin{aligned} d_\text{i} &= \left(\tfrac{1}{f} - \tfrac{1}{d_\text{o}}\right)^{-1} \\ &= \left(\tfrac{1}{10.0\text{ cm}} - \tfrac{1}{50.0\text{ cm}}\right)^{-1} \\ &= 12.5\text{ cm} \end{aligned}$$

The image is positive, so the image, is real, is on the opposite side of the lens from the object, and is 12.6 cm from the lens. To find the magnification and orientation of the image, use
Equation:

$$m = -\frac{d_\text{i}}{d_\text{o}} = -\frac{12.5\text{ cm}}{50.0\text{ cm}} = -0.250.$$

The negative magnification means that the image is inverted. Since $|m| < 1$, the image is smaller than the object. The size of the image is given by
Equation:

$$|h_\text{i}| = |m| h_\text{o} = (0.250)(3.0\text{ cm}) = 0.75\text{ cm}$$

b. For $d_\text{o} = 5.00\text{ cm}$, $f = +10.0\text{ cm}$
Equation:

$$\begin{aligned} d_\text{i} &= \left(\tfrac{1}{f} - \tfrac{1}{d_\text{o}}\right)^{-1} \\ &= \left(\tfrac{1}{10.0\text{ cm}} - \tfrac{1}{5.00\text{ cm}}\right)^{-1} \\ &= -10.0\text{ cm} \end{aligned}$$

The image distance is negative, so the image is virtual, is on the same side of the lens as the object, and is 10 cm from the lens. The magnification and orientation of the image are found from
Equation:

$$m = -\frac{d_\text{i}}{d_\text{o}} = -\frac{-10.0\text{ cm}}{5.00\text{ cm}} = +2.00.$$

The positive magnification means that the image is upright (i.e., it has the same orientation as the object). Since $|m| > 0$, the image is larger than the object. The size of the image is

Equation:

$$|h_\text{i}| = |m|h_\text{o} = (2.00)(3.0\text{ cm}) = \ 6.0\text{ cm}.$$

c. For $d_\text{o} = 20\text{ cm}, f \ = +10\text{ cm}$

Equation:

$$\begin{aligned} d_\text{i} & = \left(\frac{1}{f} - \frac{1}{d_\text{o}}\right)^{-1} \\ & = \left(\frac{1}{10.0\text{ cm}} - \frac{1}{20.0\text{ cm}}\right)^{-1} \\ & = 20.0\text{ cm} \end{aligned}$$

The image distance is positive, so the image is real, is on the opposite side of the lens from the object, and is 20.0 cm from the lens. The magnification is

Equation:

$$m = -\frac{d_\text{i}}{d_\text{o}} = -\frac{20.0\text{ cm}}{20.0\text{ cm}} = -1.00.$$

The negative magnification means that the image is inverted. Since $|m| = 1$, the image is the same size as the object.

When solving problems in geometric optics, we often need to combine ray tracing and the lens equations. The following example demonstrates this approach.

Example:

Choosing the Focal Length and Type of Lens

To project an image of a light bulb on a screen 1.50 m away, you need to choose what type of lens to use (converging or diverging) and its focal length ([link]). The distance between the lens and the lightbulb is fixed at 0.75 m. Also, what is the magnification and orientation of the image?

Strategy

The image must be real, so you choose to use a converging lens. The focal length can be found by using the thin-lens equation and solving for the focal length. The object distance is $d_\mathrm{o} = 0.75\ \mathrm{m}$ and the image distance is $d_\mathrm{i} = 1.5\ \mathrm{m}$.

Solution

Solve the thin lens for the focal length and insert the desired object and image distances:

Equation:

$$\begin{aligned} \frac{1}{d_\mathrm{o}} + \frac{1}{d_\mathrm{i}} &= \frac{1}{f} \\ f &= \left(\frac{1}{d_\mathrm{o}} + \frac{1}{d_\mathrm{i}}\right)^{-1} \\ &= \left(\frac{1}{0.75\ \mathrm{m}} + \frac{1}{1.5\ \mathrm{m}}\right)^{-1} \\ &= 0.50\ \mathrm{m} \end{aligned}$$

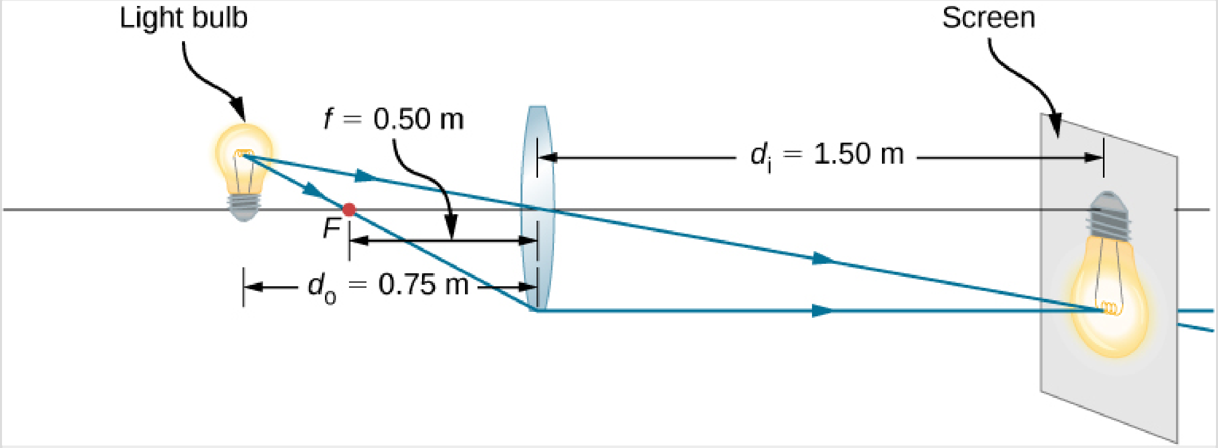
The magnification is

Equation:

$$m = -\frac{d_i}{d_\mathrm{o}} = -\frac{1.5\ \mathrm{m}}{0.75\ \mathrm{m}} = -2.0.$$

Significance

The minus sign for the magnification means that the image is inverted. The focal length is positive, as expected for a converging lens. Ray tracing can be used to check the calculation (see [link]). As expected, the image is inverted, is real, and is larger than the object.



A light bulb placed 0.75 m from a lens having a 0.50-m focal length produces a real image on a screen, as discussed in the example. Ray tracing predicts the image location and size.

Summary

- Two types of lenses are possible: converging and diverging. A lens that causes light rays to bend toward (away from) its optical axis is a converging (diverging) lens.
- For a converging lens, the focal point is where the converging light rays cross; for a diverging lens, the focal point is the point from which the diverging light rays appear to originate.
- The distance from the center of a thin lens to its focal point is called the focal length f.
- Ray tracing is a geometric technique to determine the paths taken by light rays through thin lenses.
- A real image can be projected onto a screen.
- A virtual image cannot be projected onto a screen.
- A converging lens forms either real or virtual images, depending on the object location; a diverging lens forms only virtual images.

Conceptual Questions

Exercise:

Problem:

You can argue that a flat piece of glass, such as in a window, is like a lens with an infinite focal length. If so, where does it form an image? That is, how are d_i and d_o related?

Exercise:

Problem:

When you focus a camera, you adjust the distance of the lens from the film. If the camera lens acts like a thin lens, why can it not be a fixed distance from the film for both near and distant objects?

Solution:

The focal length of the lens is fixed, so the image distance changes as a function of object distance.

Exercise:

Problem:

A thin lens has two focal points, one on either side of the lens at equal distances from its center, and should behave the same for light entering from either side. Look backward and forward through a pair of eyeglasses and comment on whether they are thin lenses.

Exercise:

Problem:

Will the focal length of a lens change when it is submerged in water? Explain.

Solution:

Yes, the focal length will change. The lens maker’s equation shows that the focal length depends on the index of refraction of the medium surrounding the lens. Because the index of refraction of water differs from that of air, the focal length of the lens will change when submerged in water.

Problems

Exercise:

Problem:

How far from the lens must the film in a camera be, if the lens has a 35.0-mm focal length and is being used to photograph a flower 75.0 cm away? Explicitly show how you follow the steps in the [link].

Exercise:

Problem:

A certain slide projector has a 100 mm-focal length lens. (a) How far away is the screen if a slide is placed 103 mm from the lens and produces a sharp image? (b) If the slide is 24.0 by 36.0 mm, what are the dimensions of the image? Explicitly show how you follow the steps in the [link].

Solution:

a. $\frac{1}{d_\mathrm{i}}+\frac{1}{d_\mathrm{o}}=\frac{1}{f}\Rightarrow d_\mathrm{i}=3.43\ \mathrm{m}$;
b. $m=-33.33$, so that

$\left(2.40\ \times\ 10^{-2}\ \text{m}\right)(33.33) = 80.0\ \text{cm, and}$
$\left(3.60\ \times\ 10^{-2}\ \text{m}\right)(33.33) = 1.20\ \text{m} \Rightarrow 0.800\ \text{m}\ \times\ 1.20\ \text{m or}\ 80.0\ \text{cm}\ \ \times\ 120\ \text{cm}$

Exercise:

Problem:

A doctor examines a mole with a 15.0-cm focal length magnifying glass held 13.5 cm from the mole. (a) Where is the image? (b) What is its magnification? (c) How big is the image of a 5.00 mm diameter mole?

Exercise:

Problem:

A camera with a 50.0-mm focal length lens is being used to photograph a person standing 3.00 m away. (a) How far from the lens must the film be? (b) If the film is 36.0 mm high, what fraction of a 1.75-m-tall person will fit on it? (c) Discuss how reasonable this seems, based on your experience in taking or posing for photographs.

Solution:

a. $\frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} = \frac{1}{f}$;
$d_\text{i} = 5.08\ \text{cm}$
b. $m = -1.695\ \times\ 10^{-2}$, so the maximum height is
$\frac{0.036\ \text{m}}{1.695 \times 10^{-2}} = 2.12\ \text{m} \Rightarrow 100\%$;
c. This seems quite reasonable, since at 3.00 m it is possible to get a full length picture of a person.

Exercise:

Problem:

A camera lens used for taking close-up photographs has a focal length of 22.0 mm. The farthest it can be placed from the film is 33.0 mm. (a) What is the closest object that can be photographed? (b) What is the magnification of this closest object?

Exercise:

Problem:

Suppose your 50.0 mm-focal length camera lens is 51.0 mm away from the film in the camera. (a) How far away is an object that is in focus? (b) What is the height of the object if its image is 2.00 cm high?

Solution:

a. $\frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} = \frac{1}{f} \Rightarrow d_\text{o} = 2.55\,\text{m}$;
b. $\frac{h_\text{i}}{h_\text{o}} = -\frac{d_\text{i}}{d_\text{o}} \Rightarrow h_\text{o} = 1.00\,\text{m}$

Exercise:

Problem:

What is the focal length of a magnifying glass that produces a magnification of 3.00 when held 5.00 cm from an object, such as a rare coin?

Exercise:

Problem:

The magnification of a book held 7.50 cm from a 10.0 cm-focal length lens is 3.00. (a) Find the magnification for the book when it is held 8.50 cm from the magnifier. (b) Repeat for the book held 9.50 cm from the magnifier. (c) Comment on how magnification changes as the object distance increases as in these two calculations.

Solution:

a. Using $\frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} = \frac{1}{f}$, $d_\text{i} = -56.67\,\text{cm}$. Then we can determine the magnification, $m = 6.67$. b. $d_\text{i} = -190\,\text{cm}$ and $m = +20.0$; c. The magnification *m* increases rapidly as you increase the object distance toward the focal length.

Exercise:

Problem:

Suppose a 200 mm-focal length telephoto lens is being used to photograph mountains 10.0 km away. (a) Where is the image? (b) What is the height of the image of a 1000 m high cliff on one of the mountains?

Exercise:

Problem:

A camera with a 100 mm-focal length lens is used to photograph the sun. What is the height of the image of the sun on the film, given the sun is 1.40×10^6 km in diameter and is 1.50×10^8 km away?

Solution:

$$\begin{aligned}\frac{1}{d_{\mathrm{o}}}+\frac{1}{d_{\mathrm{i}}} &= \frac{1}{f}\\ d_{\mathrm{i}} &= \frac{1}{(1/f)-(1/d_{\mathrm{o}})}\\ \frac{d_{\mathrm{i}}}{d_{\mathrm{o}}} &= 6.667\ \times\ 10^{-13}=\frac{h_{\mathrm{i}}}{h_{\mathrm{o}}}\\ h_{\mathrm{i}} &= -0.933\,\mathrm{mm}\end{aligned}$$

Exercise:

Problem:

Use the thin-lens equation to show that the magnification for a thin lens is determined by its focal length and the object distance and is given by $m=f/(f-d_{\mathrm{o}})$.

Exercise:

Problem:

An object of height 3.0 cm is placed 5.0 cm in front of a converging lens of focal length 20 cm and observed from the other side. Where and how large is the image?

Solution:

$d_{\mathrm{i}}=-6.7\,\mathrm{cm}$
$h_{\mathrm{i}}=4.0\,\mathrm{cm}$

Exercise:

Problem:

An object of height 3.0 cm is placed at 5.0 cm in front of a diverging lens of focal length 20 cm and observed from the other side. Where and how large is the image?

Exercise:

Problem:

An object of height 3.0 cm is placed at 25 cm in front of a diverging lens of focal length 20 cm. Behind the diverging lens, there is a converging lens of focal length 20 cm. The distance between the lenses is 5.0 cm. Find the location and size of the final image.

Solution:

83 cm to the right of the converging lens, $m=-2.3, h_{\mathrm{i}}=6.9$ cm

Exercise:

Problem:

Two convex lenses of focal lengths 20 cm and 10 cm are placed 30 cm apart, with the lens with the longer focal length on the right. An object of height 2.0 cm is placed midway between them and observed through each lens from the left and from the right. Describe what you will see, such as where the image(s) will appear, whether they will be upright or inverted and their magnifications.

Glossary

converging (or convex) lens	lens in which light rays that enter it parallel converge into a single point on the opposite side
diverging (or concave) lens	lens that causes light rays to bend away from its optical axis
focal plane	plane that contains the focal point and is perpendicular to the optical axis
ray tracing	technique that uses geometric constructions to find and characterize the image formed by an optical system
thin-lens approximation	assumption that the lens is very thin compared to the first image distance

The Eye

By the end of this section, you will be able to:

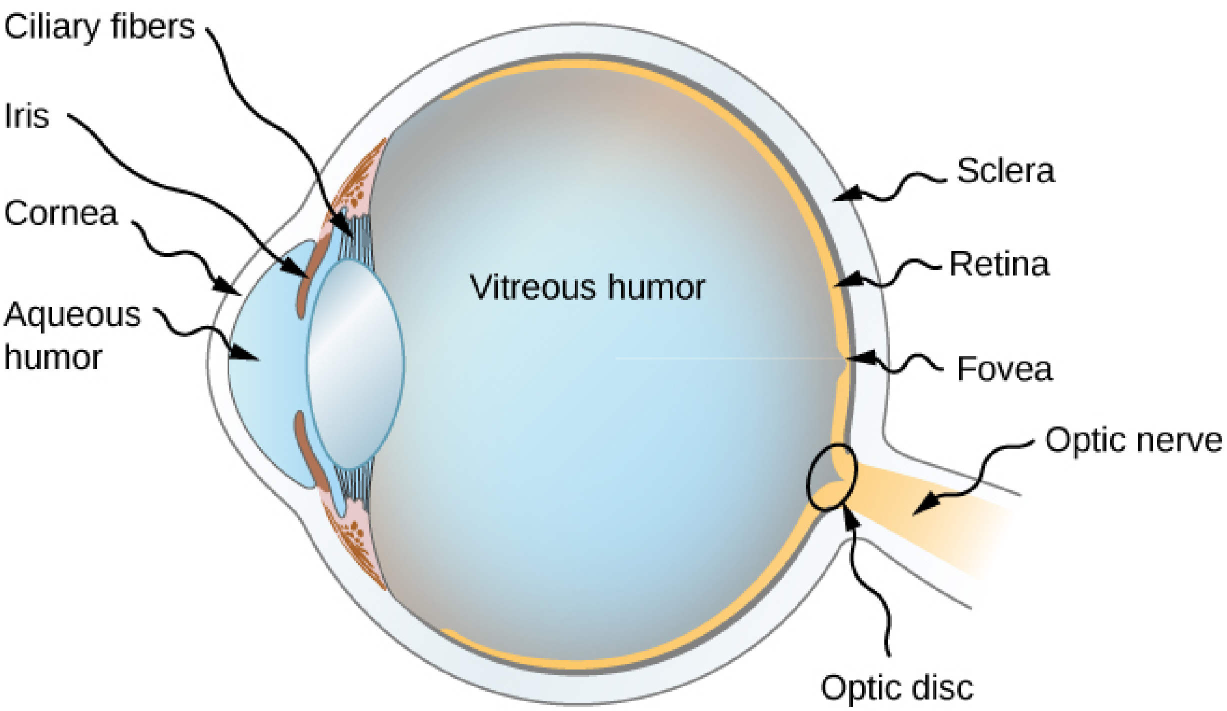
- Understand the basic physics of how images are formed by the human eye
- Recognize several conditions of impaired vision as well as the optics principles for treating these conditions

The human eye is perhaps the most interesting and important of all optical instruments. Our eyes perform a vast number of functions: They allow us to sense direction, movement, colors, and distance. In this section, we explore the geometric optics of the eye.

Physics of the Eye

The eye is remarkable in how it forms images and in the richness of detail and color it can detect. However, our eyes often need some correction to reach what is called “normal” vision. Actually, normal vision should be called “ideal” vision because nearly one-half of the human population requires some sort of eyesight correction, so requiring glasses is by no means “abnormal.” Image formation by our eyes and common vision correction can be analyzed with the optics discussed earlier in this chapter.

[link] shows the basic anatomy of the eye. The cornea and lens form a system that, to a good approximation, acts as a single thin lens. For clear vision, a real image must be projected onto the light-sensitive retina, which lies a fixed distance from the lens. The flexible lens of the eye allows it to adjust the radius of curvature of the lens to produce an image on the retina for objects at different distances. The center of the image falls on the fovea, which has the greatest density of light receptors and the greatest acuity (sharpness) in the visual field. The variable opening (i.e., the pupil) of the eye, along with chemical adaptation, allows the eye to detect light intensities from the lowest observable to 10^{10} times greater (without damage). This is an incredible range of detection. Processing of visual nerve impulses begins with interconnections in the retina and continues in the brain. The optic nerve conveys the signals received by the eye to the brain.

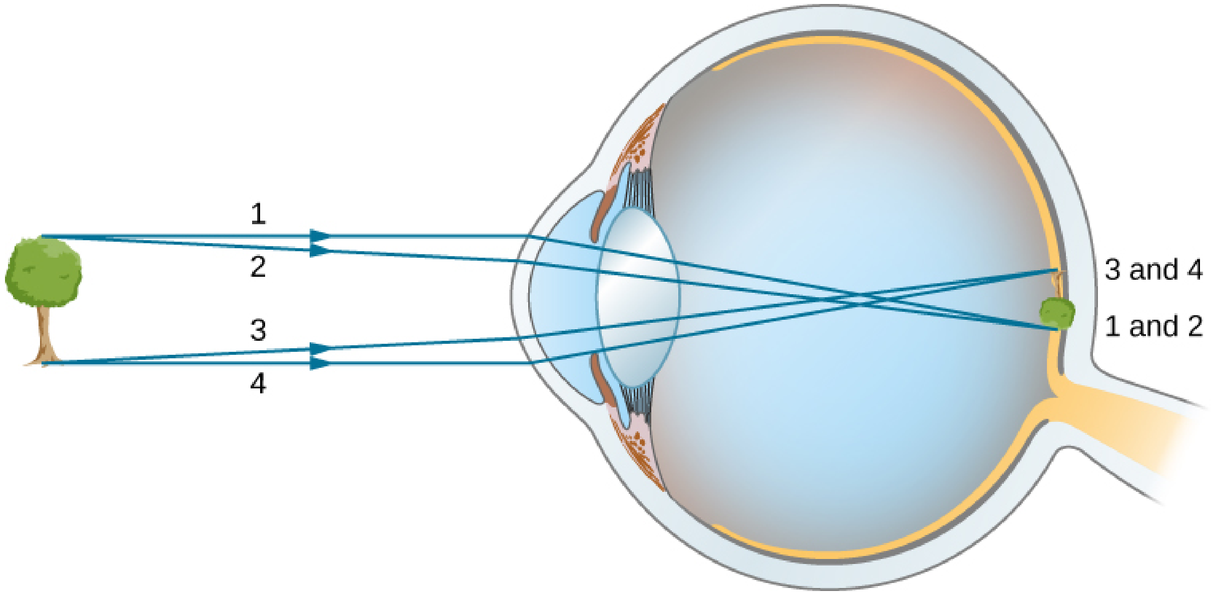


The cornea and lens of the eye act together to form a real image on the light-sensing retina, which has its densest concentration of receptors in the fovea and a blind spot over the optic nerve. The radius of curvature of the lens of an eye is adjustable to form an image on the retina for different object distances. Layers of tissues with varying indices of refraction in the lens are shown here. However, they have been omitted from other pictures for clarity.

The indices of refraction in the eye are crucial to its ability to form images. [link] lists the indices of refraction relevant to the eye. The biggest change in the index of refraction, which is where the light rays are most bent, occurs at the air-cornea interface rather than at the aqueous humor-lens interface. The ray diagram in [link] shows image formation by the cornea and lens of the eye. The cornea, which is itself a converging lens with a focal length of approximately 2.3 cm, provides most of the focusing power of the eye. The lens, which is a converging lens with a focal length of about 6.4 cm, provides the finer focus needed to produce a clear image on the retina. The cornea and lens can be treated as a single thin lens, even though the light rays pass through several layers of material (such as cornea, aqueous humor, several layers in the lens, and vitreous humor), changing direction at each interface. The image formed is much like the one produced by a single convex lens (i.e., a real, inverted image). Although images formed in the eye are inverted, the brain inverts them once more to make them seem upright.

Material	**Index of Refraction**
Water	1.33
Air	1.0
Cornea	1.38
Aqueous humor	1.34
Lens	1.41*
Vitreous humor	1.34

Refractive Indices Relevant to the Eye*This is an average value. The actual index of refraction varies throughout the lens and is greatest in center of the lens.



In the human eye, an image forms on the retina. Rays from the top and bottom of the object are traced to show how a real, inverted image is produced on the retina. The distance to the object is not to scale.

As noted, the image must fall precisely on the retina to produce clear vision—that is, the image distance d_i must equal the lens-to-retina distance. Because the lens-to-retina distance does not change, the image distance d_i must be the same for objects at all distances. The ciliary muscles adjust the shape of the eye lens for focusing on nearby or far objects. By changing the shape of the eye lens, the eye changes the focal length of the lens. This mechanism of the eye is called **accommodation**.

The nearest point an object can be placed so that the eye can form a clear image on the retina is called the **near point** of the eye. Similarly, the **far point** is the farthest distance at which an object is clearly visible. A person with normal vision can see objects clearly at distances ranging from 25 cm to essentially infinity. The near point increases with age, becoming several meters for some older people. In this text, we consider the near point to be 25 cm.

We can use the thin-lens equations to quantitatively examine image formation by the eye. First, we define the **optical power** of a lens as

Note:
Equation:

$$P = \frac{1}{f}$$

with the focal length *f* given in meters. The units of optical power are called “diopters” (D). That is, $1\,\text{D} = \frac{1}{\text{m}}$, or $1\,\text{m}^{-1}$. Optometrists prescribe common eyeglasses and contact lenses in units of diopters. With this definition of optical power, we can rewrite the thin-lens equations as

Equation:

$$P = \frac{1}{d_\text{o}} + \frac{1}{d_\text{i}}.$$

Working with optical power is convenient because, for two or more lenses close together, the effective optical power of the lens system is approximately the sum of the optical power of the individual lenses:

Note:
Equation:

$$P_\text{total} = P_\text{lens 1} + P_\text{lens 2} + P_\text{lens 3} + \cdots$$

Example:

Effective Focal Length of the Eye

The cornea and eye lens have focal lengths of 2.3 and 6.4 cm, respectively. Find the net focal length and optical power of the eye.

Strategy

The optical powers of the closely spaced lenses add, so $P_{\text{eye}} = P_{\text{cornea}} + P_{\text{lens}}$.

Solution

Writing the equation for power in terms of the focal lengths gives

Equation:

$$\frac{1}{f_{\text{eye}}} = \frac{1}{f_{\text{cornea}}} + \frac{1}{f_{\text{lens}}} = \frac{1}{2.3\ \text{cm}} + \frac{1}{6.4\ \text{cm}}.$$

Hence, the focal length of the eye (cornea and lens together) is

Equation:

$$f_{\text{eye}} = 1.69\ \text{cm}.$$

The optical power of the eye is

Equation:

$$P_{\text{eye}} = \frac{1}{f_{\text{eye}}} = \frac{1}{0.0169\ \text{m}} = 59\ \text{D}.$$

For clear vision, the image distance d_{i} must equal the lens-to-retina distance. Normal vision is possible for objects at distances $d_{\text{o}} = 25\ \text{cm}$ to infinity. The following example shows how to calculate the image distance for an object placed at the near point of the eye.

Example:

Image of an object placed at the near point

The net focal length of a particular human eye is 1.7 cm. An object is placed at the near point of the eye. How far behind the lens is a focused image formed?

Strategy

The near point is 25 cm from the eye, so the object distance is $d_\text{o} = 25\text{ cm}$. We determine the image distance from the lens equation:

Equation:

$$\frac{1}{d_\text{i}} = \frac{1}{f} - \frac{1}{d_\text{o}}.$$

Solution

Equation:

$$\begin{aligned} d_\text{i} &= \left(\frac{1}{f} - \frac{1}{d_\text{o}}\right)^{-1} \\ &= \left(\frac{1}{1.7\text{ cm}} - \frac{1}{25\text{ cm}}\right)^{-1} \\ &= 1.8\text{ cm} \end{aligned}$$

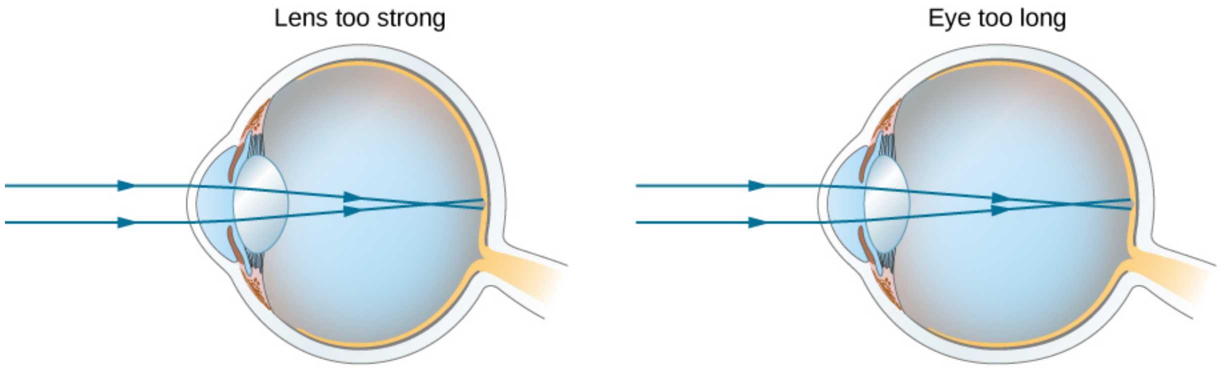
Therefore, the image is formed 1.8 cm behind the lens.

Significance

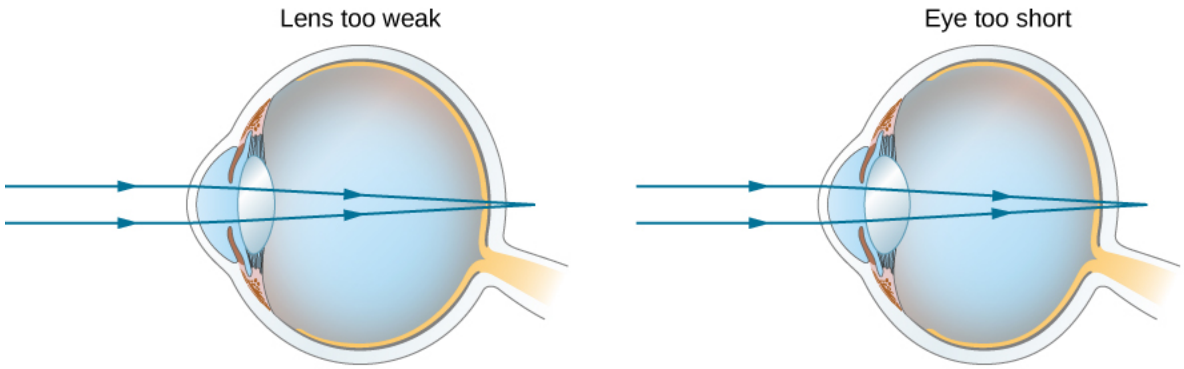
From the magnification formula, we find $m = -\frac{1.8\text{ cm}}{25\text{ cm}} = -0.073$. Since $m < 0$, the image is inverted in orientation with respect to the object. From the absolute value of *m* we see that the image is much smaller than the object; in fact, it is only 7% of the size of the object.

Vision Correction

The need for some type of vision correction is very common. Typical vision defects are easy to understand with geometric optics, and some are simple to correct. [link] illustrates two common vision defects. **Nearsightedness**, or **myopia**, is the ability to see near objects, whereas distant objects are blurry. The eye overconverges the nearly parallel rays from a distant object, and the rays cross in front of the retina. More divergent rays from a close object are converged on the retina for a clear image. The distance to the farthest object that can be seen clearly is called the far point of the eye (normally the far point is at infinity). **Farsightedness**, or **hyperopia**, is the ability to see far objects clearly, whereas near objects are blurry. A farsighted eye does not sufficiently converge the rays from a near object to make the rays meet on the retina.



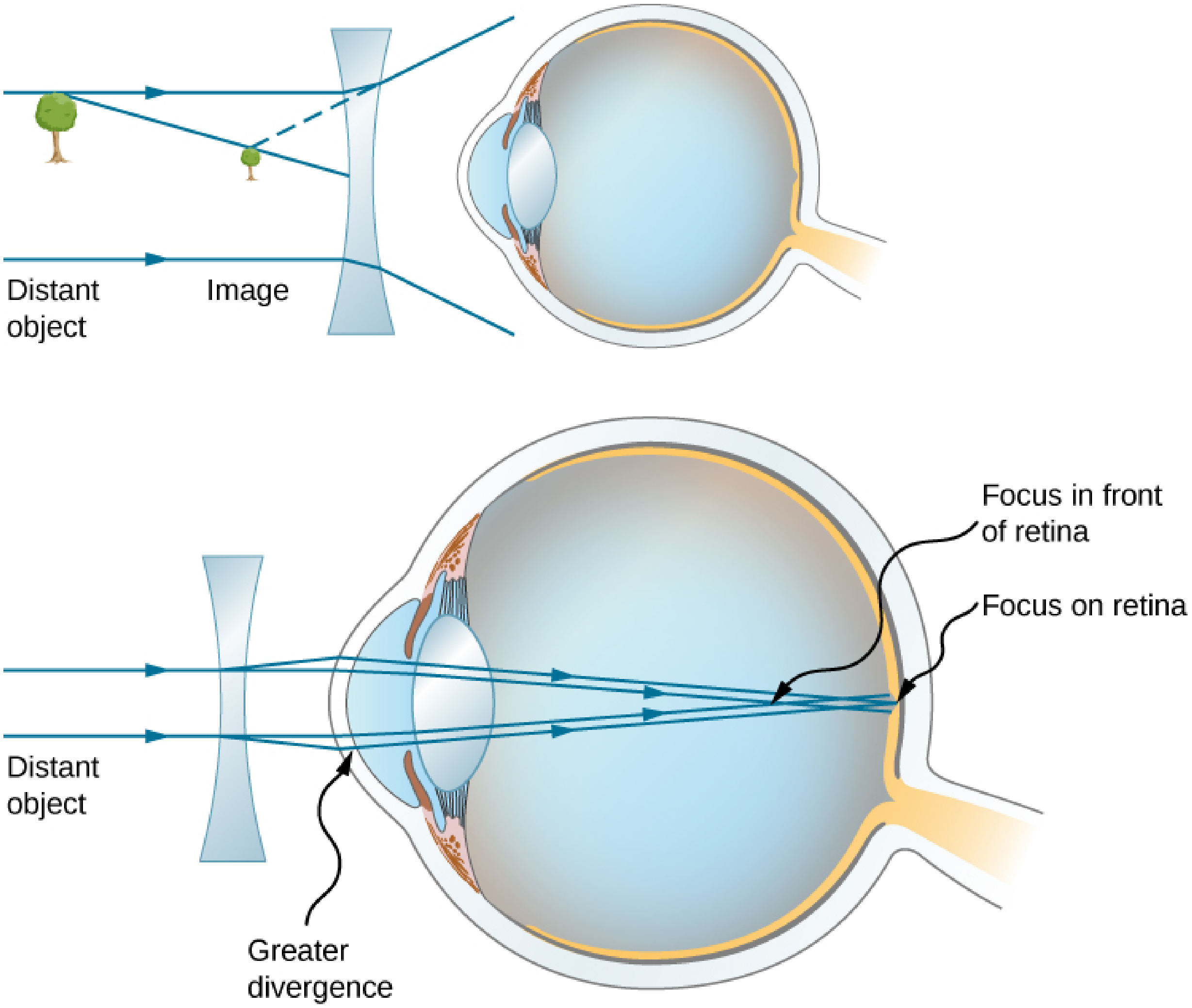
(a) Myopia



(b) Hyperopia

(a) The nearsighted (myopic) eye converges rays from a distant object in front of the retina, so they have diverged when they strike the retina, producing a blurry image. An eye lens that is too powerful can cause nearsightedness, or the eye may be too long. (b) The farsighted (hyperopic) eye is unable to converge the rays from a close object on the retina, producing blurry near-field vision. An eye lens with insufficient optical power or an eye that is too short can cause farsightedness.

Since the nearsighted eye overconverges light rays, the correction for nearsightedness consists of placing a diverging eyeglass lens in front of the eye, as shown in [link]. This reduces the optical power of an eye that is too powerful (recall that the focal length of a diverging lens is negative, so its optical power is negative). Another way to understand this correction is that a diverging lens will cause the incoming rays to diverge more to compensate for the excessive convergence caused by the lens system of the eye. The image produced by the diverging eyeglass lens serves as the (optical) object for the eye, and because the eye cannot focus on objects beyond its far point, the diverging lens must form an image of distant (physical) objects at a point that is closer than the far point.



Correction of nearsightedness requires a diverging lens that compensates for overconvergence by the eye. The diverging lens produces an image closer to the eye than the physical object. This image serves as the optical object for the eye, and the nearsighted person can see it clearly because it is closer than their far point.

Example:

Correcting Nearsightedness

What optical power of eyeglass lens is needed to correct the vision of a nearsighted person whose far point is 30.0 cm? Assume the corrective lens is fixed 1.50 cm away from the eye.

Strategy

You want this nearsighted person to be able to see distant objects clearly, which means that the eyeglass lens must produce an image 30.0 cm from the eye for an object at infinity. An image 30.0 cm from the eye will be $30.0\text{ cm}-1.50\text{ cm}=28.5\text{ cm}$ from the eyeglass lens. Therefore, we must have $d_\text{i}=-28.5\text{ cm}$ when $d_\text{o}=\infty$. The image distance is negative because it is on the same side of the eyeglass lens as the object.

Solution

Since d_i and d_o are known, we can find the optical power of the eyeglass lens by using [link]:

Equation:

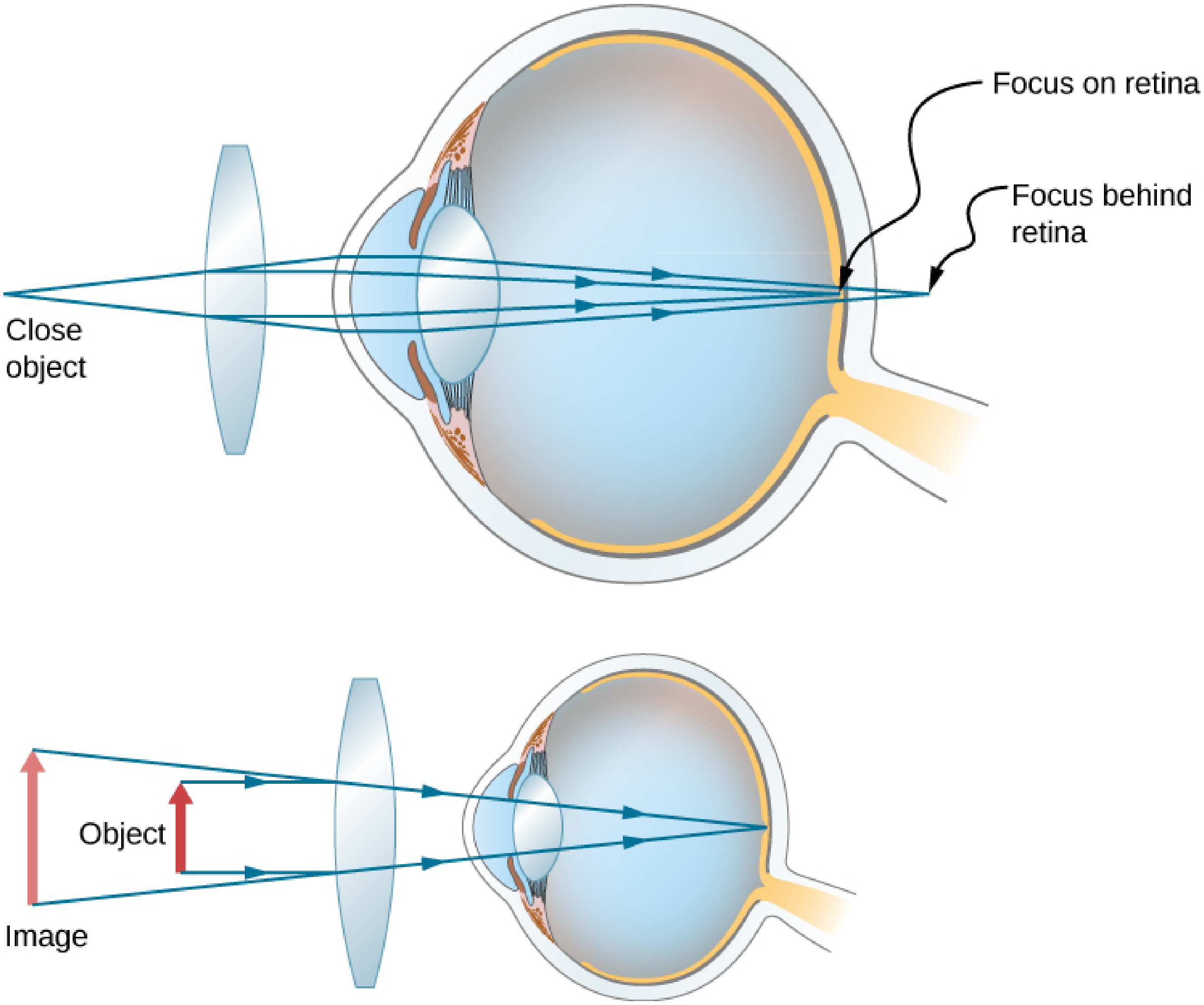
$$P=\frac{1}{d_\text{o}}+\frac{1}{d_\text{i}}=\frac{1}{\infty}+\frac{1}{-0.285\text{ m}}=-3.51\text{D}.$$

Significance

The negative optical power indicates a diverging (or concave) lens, as expected. If you examine eyeglasses for nearsighted people, you will find the lenses are thinnest in the center. Additionally, if you examine a prescription for eyeglasses for nearsighted people, you will find that the prescribed optical power is negative and given in units of diopters.

Correcting farsightedness consists simply of using the opposite type of lens as for nearsightedness (i.e., a converging lens), as shown in [link].

Such a lens will produce an image of physical objects that are closer than the near point at a distance that is between the near point and the far point, so that the person can see the image clearly. To determine the optical power needed for correction, you must therefore know the person's near point, as explained in [link].



Correction of farsightedness uses a converging lens that compensates for the underconvergence by the eye. The converging lens produces an image farther from the eye than the object, so that the farsighted person can see it clearly.

Example:

Correcting Farsightedness

What optical power of eyeglass lens is needed to allow a farsighted person, whose near point is 1.00 m, to see an object clearly that is 25.0 cm from the eye? Assume the corrective lens is fixed 1.5 cm from the eye.

Strategy

When an object is 25.0 cm from the person's eyes, the eyeglass lens must produce an image 1.00 m away (the near point), so that the person can see it clearly. An image 1.00 m from the eye will be $100\text{ cm}-1.5\text{ cm}=98.5\text{ cm}$ from the eyeglass lens because the eyeglass lens is 1.5 cm from the eye. Therefore, $d_\text{i}=-98.5\text{ cm}$, where the minus sign indicates that the image is on the same side of the lens as the object. The object is $25.0\text{ cm}-1.5\text{ cm}=23.5\text{ cm}$ from the eyeglass lens, so $d_\text{o}=23.5\text{ cm}$.

Solution

Since d_i and d_o are known, we can find the optical power of the eyeglass lens by using [link]:

Equation:

$$P=\frac{1}{d_\text{o}}+\frac{1}{d_\text{i}}=\frac{1}{0.235\text{ m}}+\frac{1}{-0.985\text{ m}}=+3.24\,\text{D}.$$

Significance

The positive optical power indicates a converging (convex) lens, as expected. If you examine eyeglasses of farsighted people, you will find the lenses to be thickest in the center. In addition, prescription eyeglasses for farsighted people have a prescribed optical power that is positive.

Summary

- Image formation by the eye is adequately described by the thin-lens equation.
- The eye produces a real image on the retina by adjusting its focal length in a process called accommodation.
- Nearsightedness, or myopia, is the inability to see far objects and is corrected with a diverging lens to reduce the optical power of the eye.
- Farsightedness, or hyperopia, is the inability to see near objects and is corrected with a converging lens to increase the optical power of the eye.
- In myopia and hyperopia, the corrective lenses produce images at distances that fall between the person's near and far points so that images can be seen clearly.

Conceptual Questions

Exercise:

Problem:

If the lens of a person's eye is removed because of cataracts (as has been done since ancient times), why would you expect an eyeglass lens of about 16 D to be prescribed?

Exercise:

Problem:

When laser light is shone into a relaxed normal-vision eye to repair a tear by spot-welding the retina to the back of the eye, the rays entering the eye must be parallel. Why?

Solution:

A relaxed, normal-vision eye will focus parallel rays of light onto the retina.

Exercise:

Problem:

Why is your vision so blurry when you open your eyes while swimming under water? How does a face mask enable clear vision?

Exercise:

Problem:

It has become common to replace the cataract-clouded lens of the eye with an internal lens. This intraocular lens can be chosen so that the person has perfect distant vision. Will the person be able to read without glasses? If the person was nearsighted, is the power of the intraocular lens greater or less than the removed lens?

Solution:

A person with an internal lens will need glasses to read because their muscles cannot distort the lens as they do with biological lenses, so they cannot focus on near objects. To correct nearsightedness, the power of the intraocular lens must be less than that of the removed lens.

Exercise:

Problem:

If the cornea is to be reshaped (this can be done surgically or with contact lenses) to correct myopia, should its curvature be made greater or smaller? Explain.

Problems

Unless otherwise stated, the lens-to-retina distance is 2.00 cm.

Exercise:

Problem:

What is the power of the eye when viewing an object 50.0 cm away?

Solution:

$$P = 52.0\,\text{D}$$

Exercise:

Problem:

Calculate the power of the eye when viewing an object 3.00 m away.

Exercise:

Problem:

The print in many books averages 3.50 mm in height. How high is the image of the print on the retina when the book is held 30.0 cm from the eye?

Solution:

$$\frac{h_\text{i}}{h_\text{o}} = -\frac{d_\text{i}}{d_\text{o}} \Rightarrow h_\text{i} = -h_\text{o}\left(\frac{d_\text{i}}{d_\text{o}}\right) = -(3.50\,\text{mm})\left(\frac{2.00\,\text{cm}}{30.0\,\text{cm}}\right) = -0.233\,\text{mm}$$

Exercise:

Problem:

Suppose a certain person’s visual acuity is such that he can see objects clearly that form an image 4.00 μm high on his retina. What is the maximum distance at which he can read the 75.0-cm-high letters on the side of an airplane?

Exercise:

Problem:

People who do very detailed work close up, such as jewelers, often can see objects clearly at much closer distance than the normal 25 cm. (a) What is the power of the eyes of a woman who can see an object clearly at a distance of only 8.00 cm? (b) What is the image size of a 1.00-mm object, such as lettering inside a ring, held at this distance? (c) What would the size of the image be if the object were held at the normal 25.0 cm distance?

Solution:

a. $P = +62.5\text{ D}$;
b. $\frac{h_\text{i}}{h_\text{o}} = -\frac{d_\text{i}}{d_\text{o}} \Rightarrow h_\text{i} = -0.250\,\text{mm}$;
c. $h_\text{i} = -0.0800\,\text{mm}$

Exercise:

Problem:

What is the far point of a person whose eyes have a relaxed power of 50.5 D?

Exercise:

Problem:

What is the near point of a person whose eyes have an accommodated power of 53.5 D?

Solution:

$$P = \frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} \Rightarrow d_\text{o} = 28.6\,\text{cm}$$

Exercise:

Problem:

(a) A laser reshaping the cornea of a myopic patient reduces the power of his eye by 9.00 D, with a $\pm 5.0\%$ uncertainty in the final correction. What is the range of diopters for eyeglass lenses that this person might need after this procedure? (b) Was the person nearsighted or farsighted before the procedure? How do you know?

Exercise:

Problem:

The power for normal close vision is 54.0 D. In a vision-correction procedure, the power of a patient's eye is increased by 3.00 D. Assuming that this produces normal close vision, what was the patient's near point before the procedure?

Solution:

Originally, the close vision was 51.0 D. Therefore,
$P = \frac{1}{d_{\text{o}}} + \frac{1}{d_{\text{i}}} \Rightarrow d_{\text{o}} = 1.00\,\text{m}$

Exercise:

Problem:

For normal distant vision, the eye has a power of 50.0 D. What was the previous far point of a patient who had laser vision correction that reduced the power of her eye by 7.00 D, producing normal distant vision?

Exercise:

Problem:

The power for normal distant vision is 50.0 D. A severely myopic patient has a far point of 5.00 cm. By how many diopters should the power of his eye be reduced in laser vision correction to obtain normal distant vision for him?

Solution:

originally, $P = 70.0\,\text{D}$; because the power for normal distant vision is 50.0 D, the power should be decreased by 20.0 D

Exercise:

Problem:

A student’s eyes, while reading the blackboard, have a power of 51.0 D. How far is the board from his eyes?

Exercise:

Problem:

The power of a physician’s eyes is 53.0 D while examining a patient. How far from her eyes is the object that is being examined?

Solution:

$P = \frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} \Rightarrow d_\text{o} = 0.333\,\text{m}$

Exercise:

Problem:

The normal power for distant vision is 50.0 D. A young woman with normal distant vision has a 10.0% ability to accommodate (that is, increase) the power of her eyes. What is the closest object she can see clearly?

Exercise:

Problem:

The far point of a myopic administrator is 50.0 cm. (a) What is the relaxed power of his eyes? (b) If he has the normal 8.00% ability to accommodate, what is the closest object he can see clearly?

Solution:

a. $P = 52.0\,\mathrm{D}$;
$P' = 56.16\,\mathrm{D}$

b.
$\frac{1}{d_\mathrm{o}} + \frac{1}{d_\mathrm{i}} = P \Rightarrow d_\mathrm{o} = 16.2\,\mathrm{cm}$

Exercise:

Problem:

A very myopic man has a far point of 20.0 cm. What power contact lens (when on the eye) will correct his distant vision?

Exercise:

Problem:

Repeat the previous problem for eyeglasses held 1.50 cm from the eyes.

Solution:

We need $d_\mathrm{i} = -18.5\,\mathrm{cm}$ when $d_\mathrm{o} = \infty$, so
$P = -5.41\,\mathrm{D}$

Exercise:

Problem:

A myopic person sees that her contact lens prescription is –4.00 D. What is her far point?

Exercise:

Problem:

Repeat the previous problem for glasses that are 1.75 cm from the eyes.

Solution:

Let x = far point

$$\Rightarrow P = \frac{1}{-(x - 0.0175\,\mathrm{m})} + \frac{1}{\infty} \Rightarrow -xP + (0.0175\,\mathrm{m})P = 1$$

$$\Rightarrow x = 26.8\,\mathrm{cm}$$

Exercise:

Problem:

The contact lens prescription for a mildly farsighted person is 0.750 D, and the person has a near point of 29.0 cm. What is the power of the tear layer between the cornea and the lens if the correction is ideal, taking the tear layer into account?

Glossary

accommodation
use of the ciliary muscles to adjust the shape of the eye lens for focusing on near or far objects

far point
furthest point an eye can see in focus

farsightedness (or hyperopia)
visual defect in which near objects appear blurred because their images are focused behind the retina rather than on the retina; a farsighted person can see far objects clearly but near objects appear blurred

near point
closest point an eye can see in focus

nearsightedness (or myopia)
visual defect in which far objects appear blurred because their images are focused in front of the retina rather than on the retina; a nearsighted person can see near objects clearly but far objects appear blurred

optical power

(P) inverse of the focal length of a lens, with the focal length expressed in meters. The optical power P of a lens is expressed in units of diopters D; that is, $1\mathrm{D} = 1/\mathrm{m} = 1\,\mathrm{m}^{-1}$

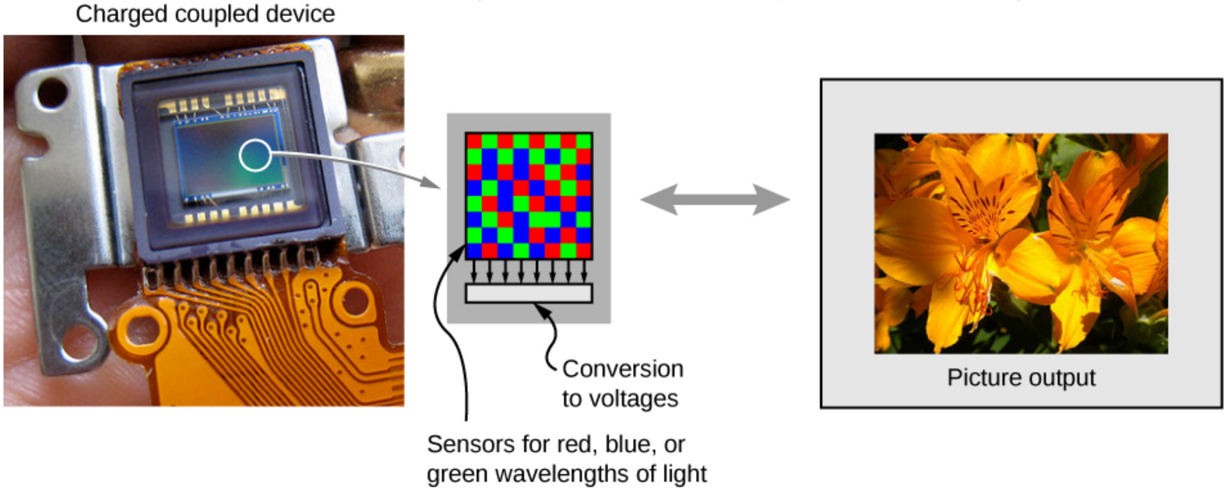
The Camera

By the end of this section, you will be able to:

- Describe the optics of a camera
- Characterize the image created by a camera

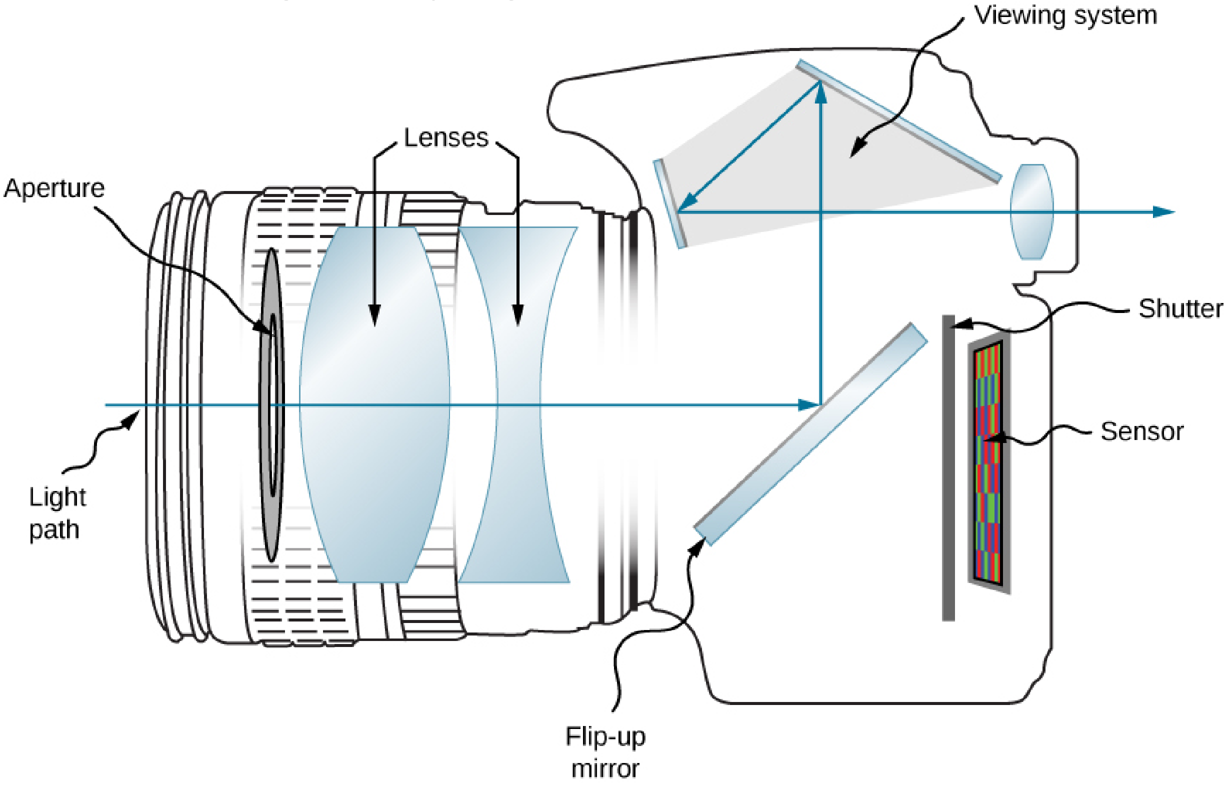
Cameras are very common in our everyday life. Between 1825 and 1827, French inventor Nicéphore Niépce successfully photographed an image created by a primitive camera. Since then, enormous progress has been achieved in the design of cameras and camera-based detectors.

Initially, photographs were recorded by using the light-sensitive reaction of silver-based compounds such as silver chloride or silver bromide. Silver-based photographic paper was in common use until the advent of digital photography in the 1980s, which is intimately connected to **charge-coupled device (CCD)** detectors. In a nutshell, a CCD is a semiconductor chip that records images as a matrix of tiny pixels, each pixel located in a “bin” in the surface. Each pixel is capable of detecting the intensity of light impinging on it. Color is brought into play by putting red-, blue-, and green-colored filters over the pixels, resulting in colored digital images ([link]). At its best resolution, one CCD pixel corresponds to one pixel of the image. To reduce the resolution and decrease the size of the file, we can “bin” several CCD pixels into one, resulting in a smaller but “pixelated” image.



A charge-coupled device (CCD) converts light signals into electronic signals, enabling electronic processing and storage of visual images. This is the basis for electronic imaging in all digital cameras, from cell phones to movie cameras. (credit left: modification of work by Bruce Turner)

Clearly, electronics is a big part of a digital camera; however, the underlying physics is basic optics. As a matter of fact, the optics of a camera are pretty much the same as those of a single lens with an object distance that is significantly larger than the lens's focal distance ([link]).



Modern digital cameras have several lenses to produce a clear image with minimal aberration and use red, blue, and green filters to produce a color image.

For instance, let us consider the camera in a smartphone. An average smartphone camera is equipped with a stationary wide-angle lens with a focal length of about 4–5 mm. (This focal length is about equal to the thickness of the phone.) The image created by the lens is focused on the

CCD detector mounted at the opposite side of the phone. In a cell phone, the lens and the CCD cannot move relative to each other. So how do we make sure that both the images of a distant and a close object are in focus?

Recall that a human eye can accommodate for distant and close images by changing its focal distance. A cell phone camera cannot do that because the distance from the lens to the detector is fixed. Here is where the small focal distance becomes important. Let us assume we have a camera with a 5-mm focal distance. What is the image distance for a selfie? The object distance for a selfie (the length of the hand holding the phone) is about 50 cm. Using the thin-lens equation, we can write

Equation:

$$\frac{1}{5\,\mathrm{mm}} = \frac{1}{500\,\mathrm{mm}} + \frac{1}{d_i}$$

We then obtain the image distance:

Equation:

$$\frac{1}{d_\mathrm{i}} = \frac{1}{5\,\mathrm{mm}} - \frac{1}{500\,\mathrm{mm}}$$

Note that the object distance is 100 times larger than the focal distance. We can clearly see that the 1/(500 mm) term is significantly smaller than 1/(5 mm), which means that the image distance is pretty much equal to the lens’s focal length. An actual calculation gives us the image distance $d_i = 5.05\,\mathrm{mm}$. This value is extremely close to the lens’s focal distance.

Now let us consider the case of a distant object. Let us say that we would like to take a picture of a person standing about 5 m from us. Using the thin-lens equation again, we obtain the image distance of 5.005 mm. The farther the object is from the lens, the closer the image distance is to the focal distance. At the limiting case of an infinitely distant object, we obtain the image distance exactly equal to the focal distance of the lens.

As you can see, the difference between the image distance for a selfie and the image distance for a distant object is just about 0.05 mm or 50 microns. Even a short object distance such as the length of your hand is two orders of magnitude larger than the lens's focal length, resulting in minute variations of the image distance. (The 50-micron difference is smaller than the thickness of an average sheet of paper.) Such a small difference can be easily accommodated by the same detector, positioned at the focal distance of the lens. Image analysis software can help improve image quality.

Conventional point-and-shoot cameras often use a movable lens to change the lens-to-image distance. Complex lenses of the more expensive mirror reflex cameras allow for superb quality photographic images. The optics of these camera lenses is beyond the scope of this textbook.

Summary

- Cameras use combinations of lenses to create an image for recording.
- Digital photography is based on charge-coupled devices (CCDs) that break an image into tiny "pixels" that can be converted into electronic signals.

Glossary

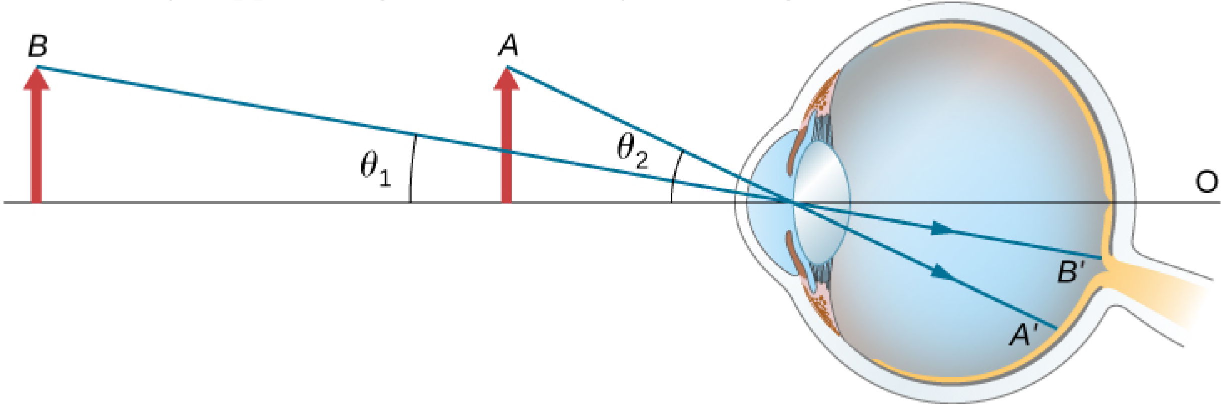
charge-coupled device (CCD)
semiconductor chip that converts a light image into tiny pixels that can be converted into electronic signals of color and intensity

The Simple Magnifier

By the end of this section, you will be able to:

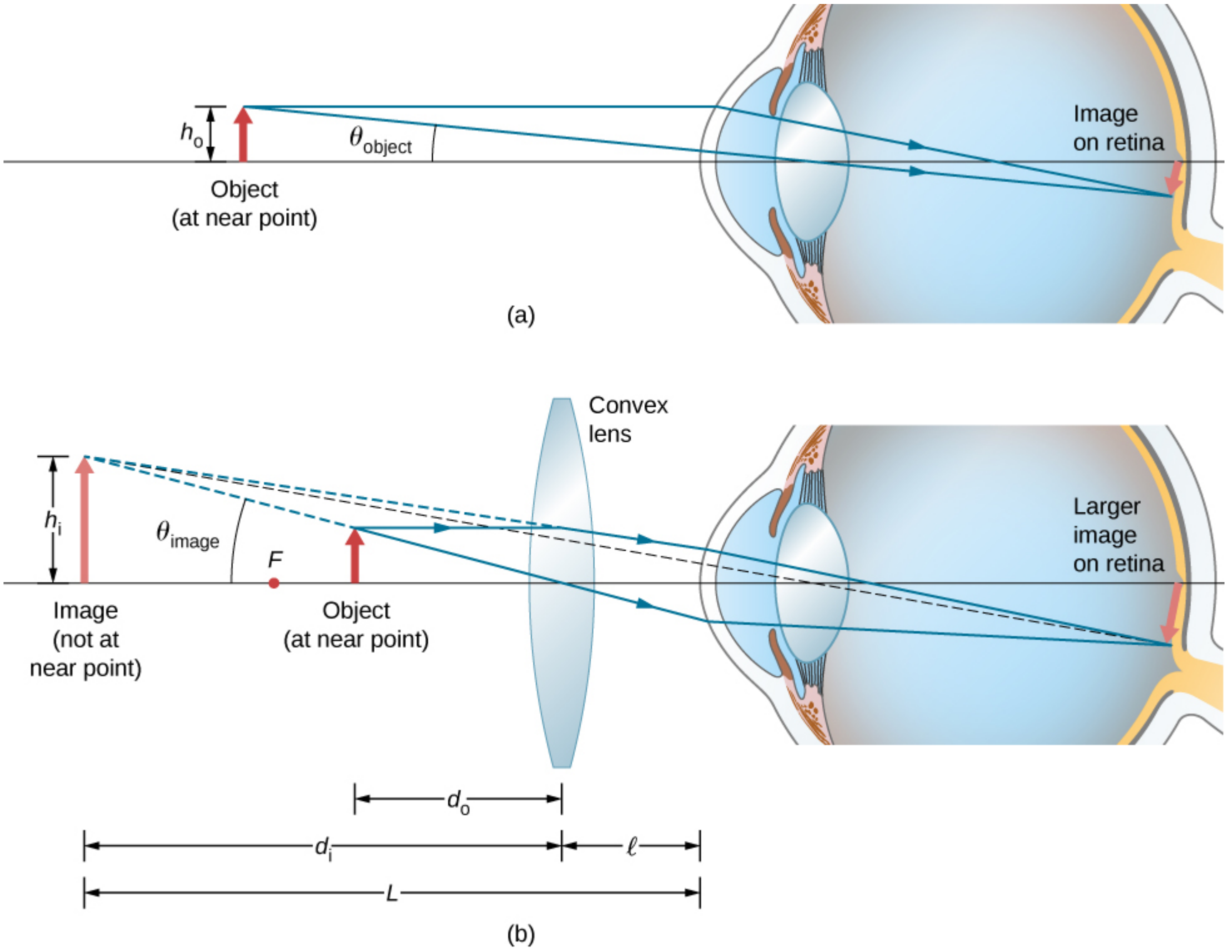
- Understand the optics of a simple magnifier
- Characterize the image created by a simple magnifier

The apparent size of an object perceived by the eye depends on the angle the object subtends from the eye. As shown in [link], the object at A subtends a larger angle from the eye than when it is position at point B. Thus, the object at A forms a larger image on the retina (see OA') than when it is positioned at B (see OB'). Thus, objects that subtend large angles from the eye appear larger because they form larger images on the retina.



Size perceived by an eye is determined by the angle subtended by the object. An image formed on the retina by an object at A is larger than an image formed on the retina by the same object positioned at B (compared image heights OA' to OB').

We have seen that, when an object is placed within a focal length of a convex lens, its image is virtual, upright, and larger than the object (see part (b) of [link]). Thus, when such an image produced by a convex lens serves as the object for the eye, as shown in [link], the image on the retina is enlarged, because the image produced by the lens subtends a larger angle in the eye than does the object. A convex lens used for this purpose is called a **magnifying glass** or a **simple magnifier**.



The simple magnifier is a convex lens used to produce an enlarged image of an object on the retina. (a) With no convex lens, the object subtends an angle θ_{object} from the eye. (b) With the convex lens in place, the image produced by the convex lens subtends an angle θ_{image} from the eye, with $\theta_{\text{image}} > \theta_{\text{object}}$. Thus, the image on the retina is larger with the convex lens in place.

To account for the magnification of a magnifying lens, we compare the angle subtended by the image (created by the lens) with the angle subtended by the object (viewed with no lens), as shown in [link]. We assume that the object is situated at the near point of the eye, because this is the object distance at which the unaided eye can form the largest image on the retina. We will compare the magnified images created by a lens with this maximum image size for the unaided eye. The magnification of an image when observed by the eye is the **angular magnification** M, which is defined by the ratio of the angle θ_{image} subtended by the image to the angle θ_{object} subtended by the object:

Note:
Equation:

$$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}}.$$

Consider the situation shown in [link]. The magnifying lens is held a distance ℓ from the eye, and the image produced by the magnifier forms a distance L from the eye. We want to calculate the angular magnification for any arbitrary L and ℓ. In the small-angle approximation, the angular size θ_{image} of the image is h_{i}/L. The angular size θ_{object} of the object at the near point is $\theta_{\text{object}} = h_{\text{o}}/25\text{ cm}$. The angular magnification is then

Equation:

$$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}} = \frac{h_{\text{i}}(25\text{ cm})}{Lh_{\text{o}}}.$$

Using [link] for linear magnification

Equation:

$$m = -\frac{d_{\text{i}}}{d_{\text{o}}} = \frac{h_{\text{i}}}{h_{\text{o}}}$$

and the thin-lens equation

Equation:

$$\frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} = \frac{1}{f}$$

in [link], we arrive at the following expression for the angular magnification of a magnifying lens:

Equation:

$$\begin{aligned} M &= \left(-\frac{d_\text{i}}{d_\text{o}}\right)\left(\frac{25\text{ cm}}{L}\right) \\ &= -d_\text{i}\left(\frac{1}{f} - \frac{1}{d_\text{i}}\right)\left(\frac{25\text{ cm}}{L}\right) \\ &= \left(1 - \frac{d_\text{i}}{f}\right)\left(\frac{25\text{ cm}}{L}\right) \end{aligned}$$

From part (b) of the figure, we see that the absolute value of the image distance is $|d_\text{i}| = L - \ell$. Note that $d_\text{i} < 0$ because the image is virtual, so we can dispense with the absolute value by explicitly inserting the minus sign: $-d_\text{i} = L - \ell$. Inserting this into [link] gives us the final equation for the angular magnification of a magnifying lens:

Note:
Equation:

$$M = \left(\frac{25\text{ cm}}{L}\right)\left(1 + \frac{L - \ell}{f}\right).$$

Note that all the quantities in this equation have to be expressed in centimeters. Often, we want the image to be at the near-point distance ($L = 25\text{ cm}$) to get maximum magnification, and we hold the magnifying lens close to the eye ($\ell = 0$). In this case, [link] gives

Equation:

$$M = 1 + \frac{25\ \text{cm}}{f}$$

which shows that the greatest magnification occurs for the lens with the shortest focal length. In addition, when the image is at the near-point distance and the lens is held close to the eye $(\ell = 0)$, then $L = d_{\text{i}} = 25\ \text{cm}$ and [link] becomes

Equation:

$$M = \frac{h_{\text{i}}}{h_{\text{o}}} = m$$

where *m* is the linear magnification ([link]) derived for spherical mirrors and thin lenses. Another useful situation is when the image is at infinity $(L = \infty)$. [link] then takes the form

Equation:

$$M(L = \infty) = \frac{25\ \text{cm}}{f}.$$

The resulting magnification is simply the ratio of the near-point distance to the focal length of the magnifying lens, so a lens with a shorter focal length gives a stronger magnification. Although this magnification is smaller by 1 than the magnification obtained with the image at the near point, it provides for the most comfortable viewing conditions, because the eye is relaxed when viewing a distant object.

By comparing [link] with [link], we see that the range of angular magnification of a given converging lens is

Note:

Equation:

$$\frac{25\text{ cm}}{f} \leq M \leq 1 + \frac{25\text{ cm}}{f}.$$

Example:

Magnifying a Diamond

A jeweler wishes to inspect a 3.0-mm-diameter diamond with a magnifier. The diamond is held at the jeweler's near point (25 cm), and the jeweler holds the magnifying lens close to his eye.

(a) What should the focal length of the magnifying lens be to see a 15-mm-diameter image of the diamond?

(b) What should the focal length of the magnifying lens be to obtain $10\ \times$ magnification?

Strategy

We need to determine the requisite magnification of the magnifier. Because the jeweler holds the magnifying lens close to his eye, we can use [link] to find the focal length of the magnifying lens.

Solution

a. The required linear magnification is the ratio of the desired image diameter to the diamond's actual diameter ([link]). Because the jeweler holds the magnifying lens close to his eye and the image forms at his near point, the linear magnification is the same as the angular magnification, so

   Equation:

$$M = m = \frac{h_\text{i}}{h_\text{o}} = \frac{15\text{ mm}}{3.0\text{ mm}} = 5.0.$$

   The focal length *f* of the magnifying lens may be calculated by solving [link] for *f*, which gives

   Equation:

$$\begin{aligned} M &= 1 + \frac{25\text{ cm}}{f} \\ f &= \frac{25\text{ cm}}{M-1} = \frac{25\text{ cm}}{5.0-1} = 6.3\text{ cm} \end{aligned}$$

b. To get an image magnified by a factor of ten, we again solve [link] for *f*, but this time we use $M = 10$. The result is
Equation:

$$f = \frac{25\text{ cm}}{M-1} = \frac{25\text{ cm}}{10-1} = 2.8\text{ cm}.$$

Significance
Note that a greater magnification is achieved by using a lens with a smaller focal length. We thus need to use a lens with radii of curvature that are less than a few centimeters and hold it very close to our eye. This is not very convenient. A compound microscope, explored in the following section, can overcome this drawback.

Summary

- A simple magnifier is a converging lens and produces a magnified virtual image of an object located within the focal length of the lens.
- Angular magnification accounts for magnification of an image created by a magnifier. It is equal to the ratio of the angle subtended by the image to that subtended by the object when the object is observed by the unaided eye.
- Angular magnification is greater for magnifying lenses with smaller focal lengths.
- Simple magnifiers can produce as great as tenfold ($10\ \times$) magnification.

Problems

Exercise:

Problem:

If the image formed on the retina subtends an angle of 30° and the object subtends an angle of 5°, what is the magnification of the image?

Solution:

$M = 6\ \times$

Exercise:

Problem:

What is the magnification of a magnifying lens with a focal length of 10 cm if it is held 3.0 cm from the eye and the object is 12 cm from the eye?

Exercise:

Problem:

How far should you hold a 2.1 cm-focal length magnifying glass from an object to obtain a magnification of $10\ \times$? Assume you place your eye 5.0 cm from the magnifying glass.

Solution:

$$\begin{aligned} M &= \left(\frac{25\text{ cm}}{L}\right)\left(1+\frac{L-\ell}{f}\right) \\ L-\ell &= d_\text{o} \\ d_\text{o} &= 13\text{ cm} \end{aligned}$$

Exercise:

Problem:

You hold a 5.0 cm-focal length magnifying glass as close as possible to your eye. If you have a normal near point, what is the magnification?

Exercise:

Problem:

You view a mountain with a magnifying glass of focal length $f = 10\ \text{cm}$. What is the magnification?

Solution:

$M = 2.5\ \times$

Exercise:

Problem:

You view an object by holding a 2.5 cm-focal length magnifying glass 10 cm away from it. How far from your eye should you hold the magnifying glass to obtain a magnification of $10\times$?

Exercise:

Problem:

A magnifying glass forms an image 10 cm on the opposite side of the lens from the object, which is 10 cm away. What is the magnification of this lens for a person with a normal near point if their eye 12 cm from the object?

Solution:

$M \quad = -2.1\times$

Exercise:

Problem:

An object viewed with the naked eye subtends a 2° angle. If you view the object through a $10\ \times$ magnifying glass, what angle is subtended by the image formed on your retina?

Exercise:

Problem:

For a normal, relaxed eye, a magnifying glass produces an angular magnification of 4.0. What is the largest magnification possible with this magnifying glass?

Solution:

$$M = \frac{25\text{ cm}}{f}$$
$$M_{\text{max}} = 5$$

Exercise:

Problem:

What range of magnification is possible with a 7.0 cm-focal length converging lens?

Exercise:

Problem:

A magnifying glass produces an angular magnification of 4.5 when used by a young person with a near point of 18 cm. What is the maximum angular magnification obtained by an older person with a near point of 45 cm?

Solution:

$$\begin{aligned} M_{\text{max}}^{\text{young}} &= 1 + \frac{18\text{ cm}}{f} \Rightarrow f = \frac{18\text{ cm}}{M_{\text{max}}^{\text{young}} - 1} \\ M_{\text{max}}^{\text{old}} &= 9.8\,\times \end{aligned}$$

Glossary

angular magnification
: ratio of the angle subtended by an object observed with a magnifier to that observed by the naked eye

simple magnifier (or magnifying glass)

converging lens that produces a virtual image of an object that is within the focal length of the lens

Microscopes and Telescopes

By the end of this section, you will be able to:

- Explain the physics behind the operation of microscopes and telescopes
- Describe the image created by these instruments and calculate their magnifications

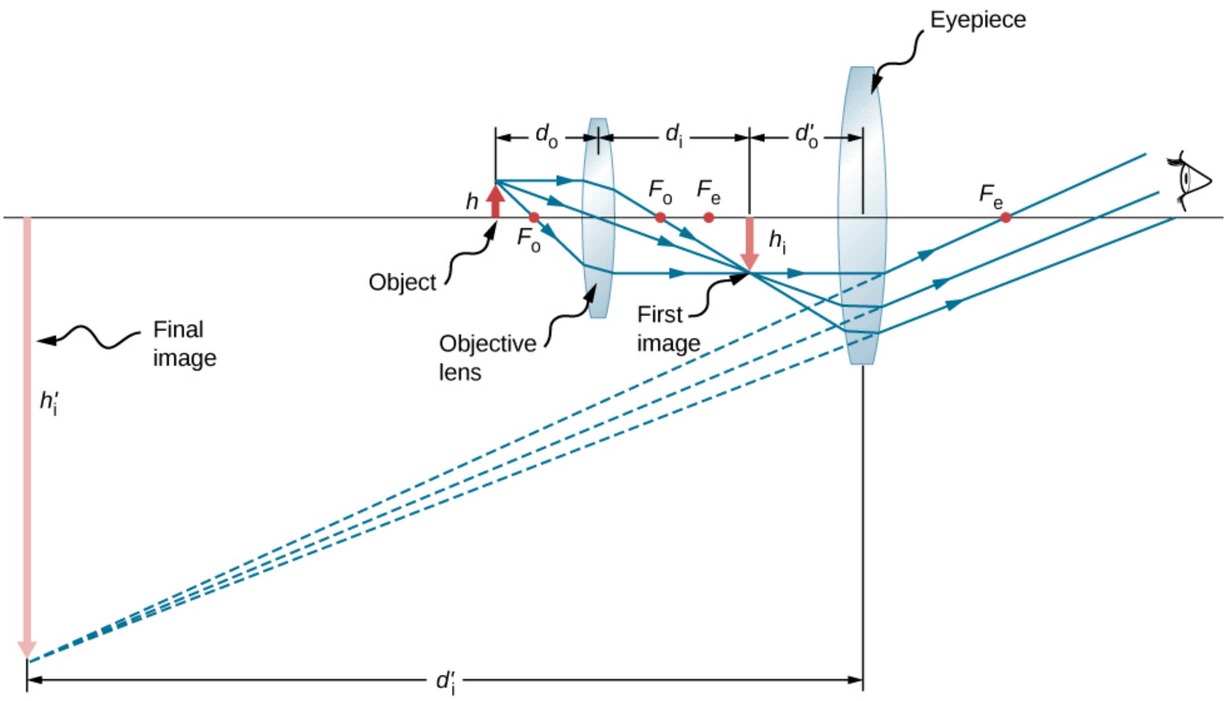
Microscopes and telescopes are major instruments that have contributed hugely to our current understanding of the micro- and macroscopic worlds. The invention of these devices led to numerous discoveries in disciplines such as physics, astronomy, and biology, to name a few. In this section, we explain the basic physics that make these instruments work.

Microscopes

Although the eye is marvelous in its ability to see objects large and small, it obviously is limited in the smallest details it can detect. The desire to see beyond what is possible with the naked eye led to the use of optical instruments. We have seen that a simple convex lens can create a magnified image, but it is hard to get large magnification with such a lens. A magnification greater than $5\times$ is difficult without distorting the image. To get higher magnification, we can combine the simple magnifying glass with one or more additional lenses. In this section, we examine microscopes that enlarge the details that we cannot see with the naked eye.

Microscopes were first developed in the early 1600s by eyeglass makers in The Netherlands and Denmark. The simplest **compound microscope** is constructed from two convex lenses ([link]). The **objective** lens is a convex lens of short focal length (i.e., high power) with typical magnification from $5\times$ to $100\times$. The **eyepiece**, also referred to as the ocular, is a convex lens of longer focal length.

The purpose of a microscope is to create magnified images of small objects, and both lenses contribute to the final magnification. Also, the final enlarged image is produced sufficiently far from the observer to be easily viewed, since the eye cannot focus on objects or images that are too close (i.e., closer than the near point of the eye).



A compound microscope is composed of two lenses: an objective and an eyepiece. The objective forms the first image, which is larger than the object. This first image is inside the focal length of the eyepiece and serves as the object for the eyepiece. The eyepiece forms final image that is further magnified.

To see how the microscope in [link] forms an image, consider its two lenses in succession. The object is just beyond the focal length f^{obj} of the objective lens, producing a real, inverted image that is larger than the object. This first image serves as the object for the second lens, or eyepiece. The eyepiece is positioned so that the first image is within its focal length f^{eye}, so that it can further magnify the image. In a sense, it acts as a magnifying glass that magnifies the intermediate image produced by the objective. The image produced by the eyepiece is a magnified virtual image. The final image remains inverted but is farther from the observer than the object, making it easy to view.

The eye views the virtual image created by the eyepiece, which serves as the object for the lens in the eye. The virtual image formed by the eyepiece is well outside the focal length of the eye, so the eye forms a real image on the retina.

The magnification of the microscope is the product of the linear magnification m^{obj} by the objective and the angular magnification M^{eye} by the eyepiece. These are given by

Equation:

$$\begin{aligned} m^{\text{obj}} &= -\frac{d_{\text{i}}^{\text{obj}}}{d_{\text{o}}^{\text{obj}}} \approx -\frac{d_{\text{i}}^{\text{obj}}}{f^{\text{obj}}} \text{ (linear magnification by objective)} \\ M^{\text{eye}} &= 1 + \frac{25\text{ cm}}{f^{\text{eye}}} \text{ (angular magnification by eyepiece)} \end{aligned}$$

Here, f^{obj} and f^{eye} are the focal lengths of the objective and the eyepiece, respectively. We assume that the final image is formed at the near point of the eye, providing the largest magnification. Note that the angular magnification of the eyepiece is the same as obtained earlier for the simple magnifying glass. This should not be surprising, because the eyepiece is essentially a magnifying glass, and the same physics applies here. The **net magnification** M_{net} of the compound microscope is the product of the linear magnification of the objective and the angular magnification of the eyepiece:

Note:

Equation:

$$M_{\text{net}} = m^{\text{obj}} M^{\text{eye}} = -\frac{d_{\text{i}}^{\text{obj}} \left(f^{\text{eye}} + 25\text{ cm}\right)}{f^{\text{obj}} f^{\text{eye}}}.$$

Example:

Microscope Magnification

Calculate the magnification of an object placed 6.20 mm from a compound microscope that has a 6.00 mm-focal length objective and a 50.0 mm-focal length eyepiece. The objective and eyepiece are separated by 23.0 cm.

Strategy

This situation is similar to that shown in [link]. To find the overall magnification, we must know the linear magnification of the objective and the angular magnification of the eyepiece. We can use [link], but we need to use the thin-lens equation to find the image distance $d_{\text{i}}^{\text{obj}}$ of the objective.

Solution

Solving the thin-lens equation for $d_{\text{i}}^{\text{obj}}$ gives

Equation:

$$\begin{aligned} d_{\text{i}}^{\text{obj}} &= \left(\frac{1}{f^{\text{obj}}} - \frac{1}{d_{\text{o}}^{\text{obj}}}\right)^{-1} \\ &= \left(\frac{1}{6.00\ \text{mm}} - \frac{1}{6.20\ \text{mm}}\right)^{-1} = 186\ \text{mm} = 18.6\ \text{cm} \end{aligned}$$

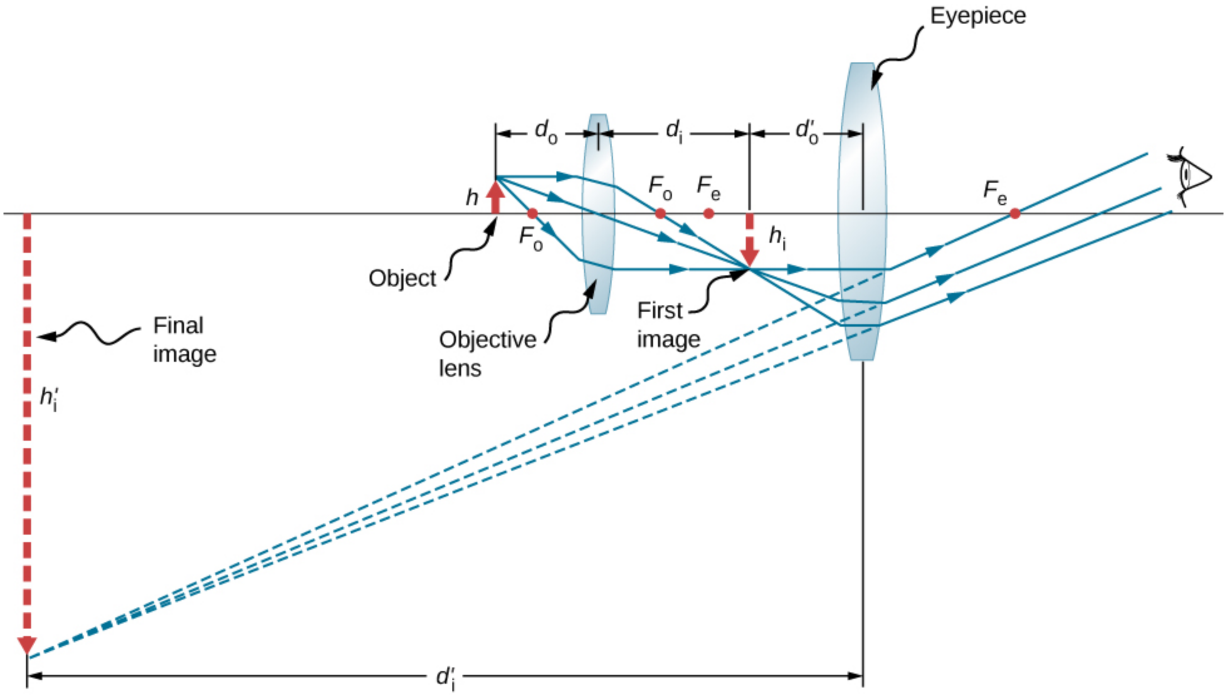
Inserting this result into [link] along with the known values $f^{\text{obj}} = 6.00\ \text{mm} = 0.600\ \text{cm}$ and $f^{\text{eye}} = 50.0\ \text{mm} = 5.00\ \text{cm}$ gives

Equation:

$$\begin{aligned} M_{\text{net}} &= -\frac{d_{\text{i}}^{\text{obj}}(f^{\text{eye}} + 25\ \text{cm})}{f^{\text{obj}} f^{\text{eye}}} \\ &= -\frac{(18.6\ \text{cm})(5.00\ \text{cm} + 25\ \text{cm})}{(0.600\ \text{cm})(5.00\ \text{cm})} \\ &= -186 \end{aligned}$$

Significance

Both the objective and the eyepiece contribute to the overall magnification, which is large and negative, consistent with [link], where the image is seen to be large and inverted. In this case, the image is virtual and inverted, which cannot happen for a single element (see [link]).



A compound microscope with the image created at infinity.

We now calculate the magnifying power of a microscope when the image is at infinity, as shown in [link], because this makes for the most relaxed viewing. The magnifying power of the microscope is the product of linear magnification m^{obj} of the objective and the angular magnification M^{eye} of the eyepiece. We know that $m^{\text{obj}} = -d_{\text{i}}^{\text{obj}}/d_{\text{o}}^{\text{obj}}$ and from the thin-lens equation we obtain

Equation:

$$m^{\text{obj}} = -\frac{d_{\text{i}}^{\text{obj}}}{d_{\text{o}}^{\text{obj}}} = 1 - \frac{d_{\text{i}}^{\text{obj}}}{f^{\text{obj}}} = \frac{f^{\text{obj}} - d_{\text{i}}^{\text{obj}}}{f^{\text{obj}}}.$$

If the final image is at infinity, then the image created by the objective must be located at the focal point of the eyepiece. This may be seen by considering the thin-lens equation with $d_{\text{i}} = \infty$ or by recalling that rays that pass through the focal point exit the lens parallel to each other, which is equivalent to focusing at infinity. For many microscopes, the distance between the image-side focal point of the objective and the object-side focal point of the eyepiece is standardized at $L=16\text{ cm}$. This distance is called the tube length of the microscope. From [link], we see that $L=f^{\text{obj}}-d_{\text{i}}^{\text{obj}}$. Inserting this into [link] gives

Equation:

$$m^{\text{obj}}=\frac{L}{f^{\text{obj}}}=\frac{16\text{ cm}}{f^{\text{obj}}}.$$

We now need to calculate the angular magnification of the eyepiece with the image at infinity. To do so, we take the ratio of the angle θ_{image} subtended by the image to the angle θ_{object} subtended by the object at the near point of the eye (this is the closest that the unaided eye can view the object, and thus this is the position where the object will form the largest image on the retina of the unaided eye). Using [link] and working in the small-angle approximation, we have $\theta_{\text{image}}\approx h_{\text{i}}^{\text{obj}}/f^{\text{eye}}$ and $\theta_{\text{object}}\approx h_{\text{i}}^{\text{obj}}/25\text{ cm}$, where $h_{\text{i}}^{\text{obj}}$ is the height of the image formed by the objective, which is the object of the eyepiece. Thus, the angular magnification of the eyepiece is

Equation:

$$M^{\text{eye}}=\frac{\theta_{\text{image}}}{\theta_{\text{object}}}=\frac{h_{\text{i}}^{\text{obj}}}{f^{\text{eye}}}\frac{25\text{ cm}}{h_{\text{i}}^{\text{obj}}}=\frac{25\text{ cm}}{f^{\text{eye}}}.$$

The net magnifying power of the compound microscope with the image at infinity is therefore

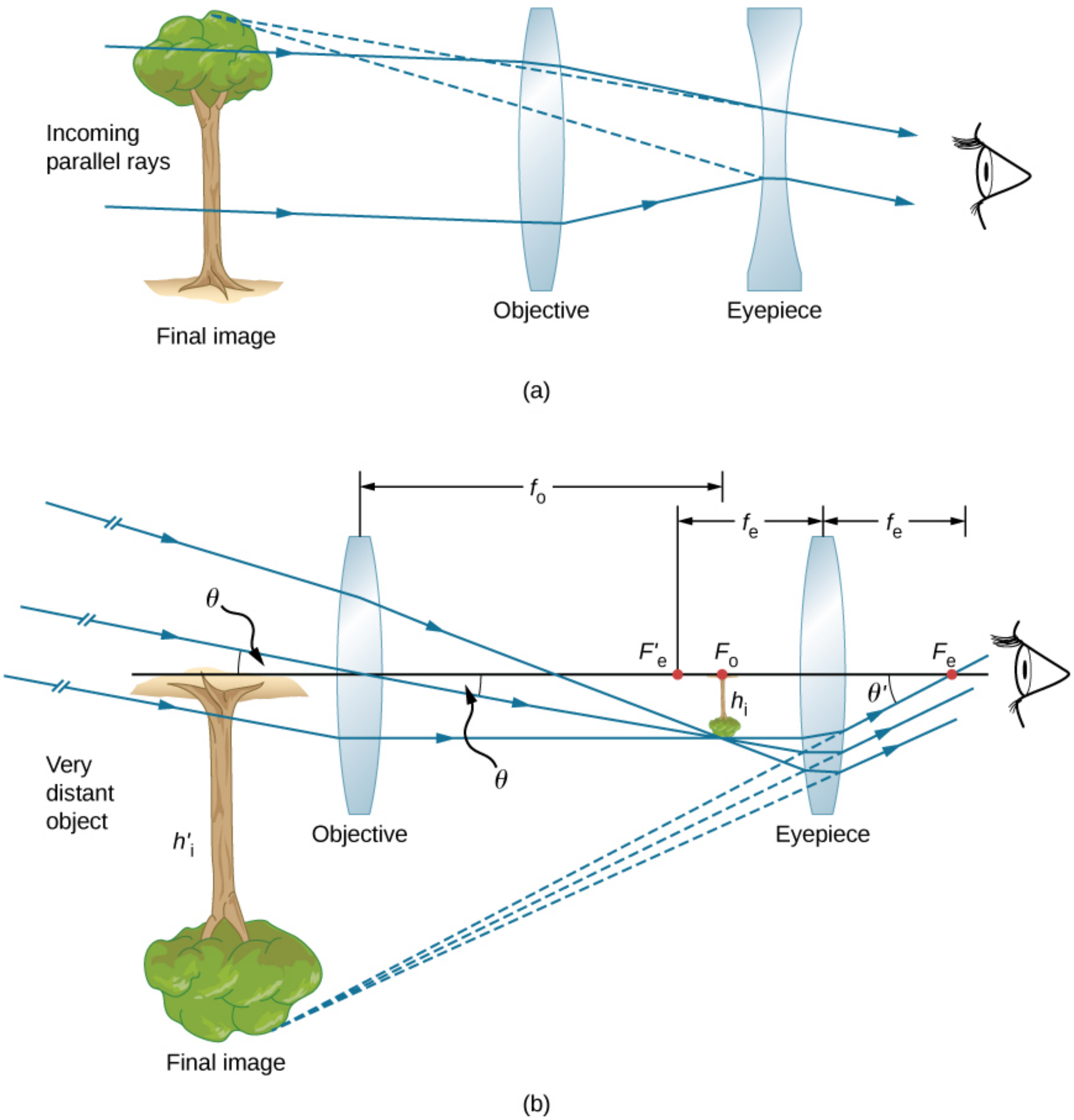
Equation:

$$M_{\text{net}}=m^{\text{obj}}M^{\text{eye}}=-\frac{(16\text{ cm})(25\text{ cm})}{f^{\text{obj}}f^{\text{eye}}}.$$

The focal distances must be in centimeters. The minus sign indicates that the final image is inverted. Note that the only variables in the equation are the focal distances of the eyepiece and the objective, which makes this equation particularly useful.

Telescopes

Telescopes are meant for viewing distant objects and produce an image that is larger than the image produced in the unaided eye. Telescopes gather far more light than the eye, allowing dim objects to be observed with greater magnification and better resolution. Telescopes were invented around 1600, and Galileo was the first to use them to study the heavens, with monumental consequences. He observed the moons of Jupiter, the craters and mountains on the moon, the details of sunspots, and the fact that the Milky Way is composed of a vast number of individual stars.



(a) Galileo made telescopes with a convex objective and a concave eyepiece. These produce an upright image and are used in spyglasses. (b) Most simple refracting telescopes have two convex lenses. The objective forms a real, inverted image at (or just within) the focal plane of the eyepiece. This image serves as the object for the eyepiece. The eyepiece forms a virtual, inverted image that is magnified.

Part (a) of [link] shows a refracting telescope made of two lenses. The first lens, called the objective, forms a real image within the focal length of the second lens, which is called the eyepiece. The image of the objective lens serves as the object for the eyepiece, which forms a magnified virtual image that is observed by the eye. This design is what Galileo used to observe the heavens.

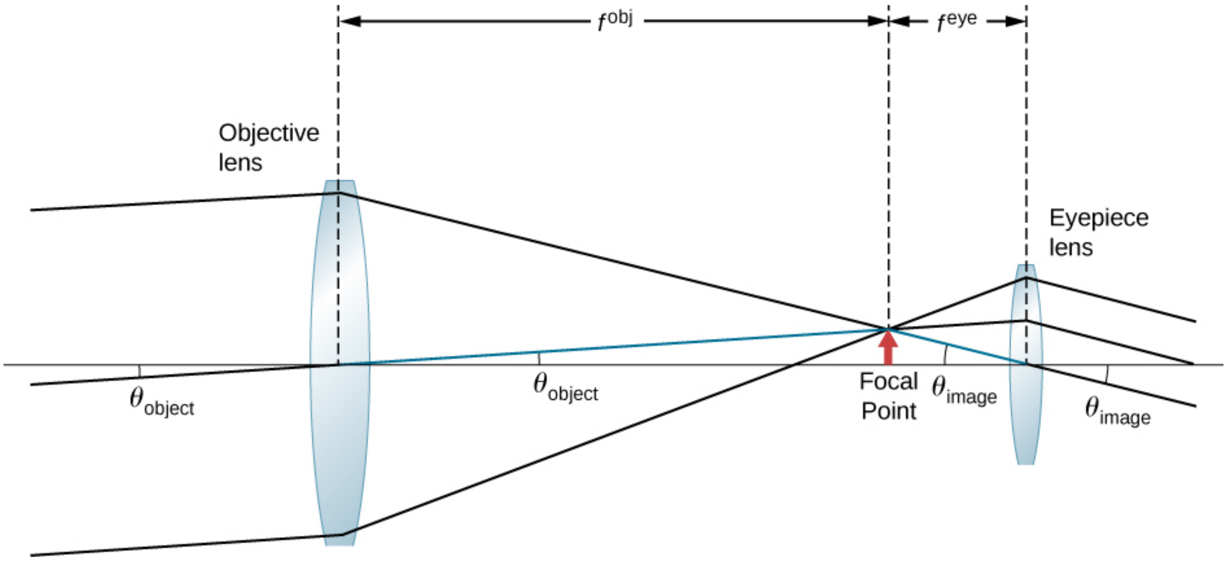
Although the arrangement of the lenses in a refracting telescope looks similar to that in a microscope, there are important differences. In a telescope, the real object is far away and the intermediate image is smaller than the object. In a microscope, the real object is very close and the intermediate image is larger than the object. In both the telescope and the microscope, the eyepiece magnifies the intermediate image; in the telescope, however, this is the only magnification.

The most common two-lens telescope is shown in part (b) of the figure. The object is so far from the telescope that it is essentially at infinity compared with the focal lengths of the lenses $\left(d_{\mathrm{o}}^{\mathrm{obj}} \approx \infty\right)$, so the incoming rays are essentially parallel and focus on the focal plane. Thus, the first image is produced at $d_{\mathrm{i}}^{\mathrm{obj}} = f^{\mathrm{obj}}$, as shown in the figure, and is not large compared with what you might see by looking directly at the object. However, the eyepiece of the telescope eyepiece (like the microscope eyepiece) allows you to get nearer than your near point to this first image and so magnifies it (because you are near to it, it subtends a larger angle from your eye and so forms a larger image on your retina). As for a simple magnifier, the angular magnification of a telescope is the ratio of the angle subtended by the image [θ_{image} in part (b)] to the angle subtended by the real object [θ_{object} in part (b)]:

Equation:

$$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}}.$$

To obtain an expression for the magnification that involves only the lens parameters, note that the focal plane of the objective lens lies very close to the focal plan of the eyepiece. If we assume that these planes are superposed, we have the situation shown in [link].



The focal plane of the objective lens of a telescope is very near to the focal plane of the eyepiece. The angle θ_{image} subtended by the image viewed through the eyepiece is larger than the angle θ_{object} subtended by the object when viewed with the unaided eye.

We further assume that the angles θ_{object} and θ_{image} are small, so that the small-angle approximation holds ($\tan\theta \approx \theta$). If the image formed at the focal plane has height *h*, then

Equation:

$$\theta_{\text{object}} \approx \tan\theta_{\text{object}} = \frac{h}{f^{\text{obj}}}$$

$$\theta_{\text{image}} \approx \tan\theta_{\text{image}} = \frac{-h}{f^{\text{eye}}}$$

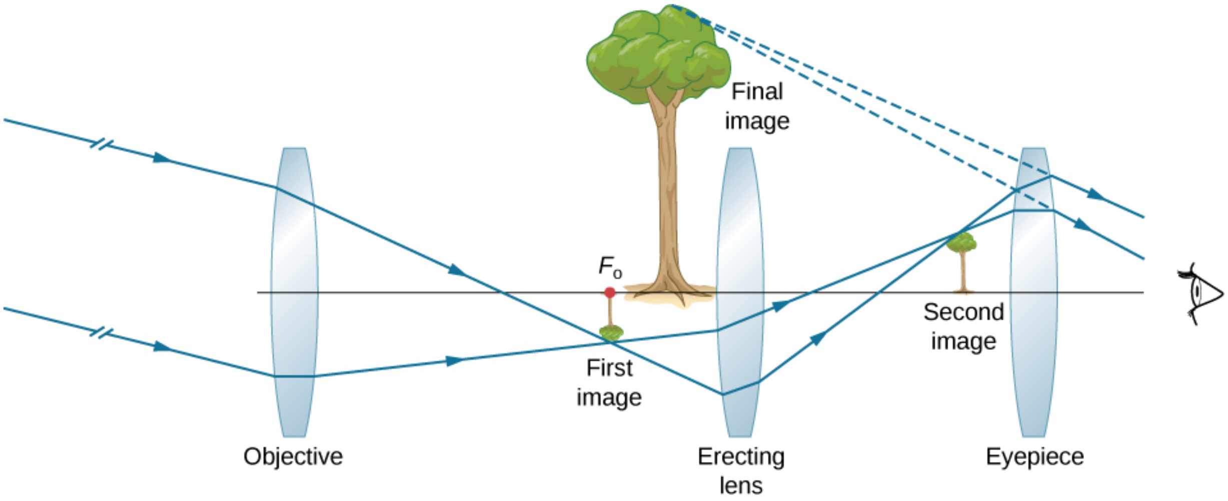
where the minus sign is introduced because the height is negative if we measure both angles in the counterclockwise direction. Inserting these expressions into [link] gives

Equation:

$$M = \frac{-h_{\text{i}}}{f^{\text{eye}}} \frac{f^{\text{obj}}}{h_{\text{i}}} = -\frac{f^{\text{obj}}}{f^{\text{eye}}}.$$

Thus, to obtain the greatest angular magnification, it is best to have an objective with a long focal length and an eyepiece with a short focal length. The greater the angular magnification *M*, the larger an object will appear when viewed through a telescope, making more details visible. Limits to observable details are imposed by many factors, including lens quality and atmospheric disturbance. Typical eyepieces have focal lengths of 2.5 cm or 1.25 cm. If the objective of the telescope has a focal length of 1 meter, then these eyepieces result in magnifications of 40 × and 80 × , respectively. Thus, the angular magnifications make the image appear 40 times or 80 times closer than the real object.

The minus sign in the magnification indicates the image is inverted, which is unimportant for observing the stars but is a real problem for other applications, such as telescopes on ships or telescopic gun sights. If an upright image is needed, Galileo’s arrangement in part (a) of [link] can be used. But a more common arrangement is to use a third convex lens as an eyepiece, increasing the distance between the first two and inverting the image once again, as seen in [link].



This arrangement of three lenses in a telescope produces an upright final image. The first two lenses are far enough apart that the second lens inverts the image of the first. The third lens acts as a magnifier and keeps the image upright and in a location that is easy to view.

The largest refracting telescope in the world is the 40-inch diameter Yerkes telescope located at Lake Geneva, Wisconsin ([link]), and operated by the University of Chicago.

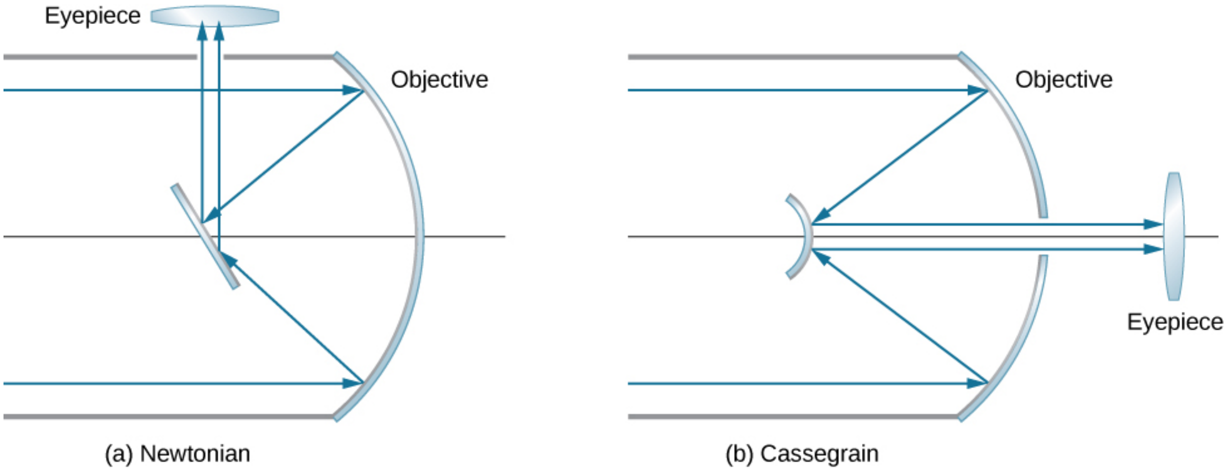
It is very difficult and expensive to build large refracting telescopes. You need large defect-free lenses, which in itself is a technically demanding task. A refracting telescope basically looks like a tube with a support structure to rotate it in different directions. A refracting telescope suffers from several problems. The aberration of lenses causes the image to be blurred. Also, as the lenses become thicker for larger lenses, more light is absorbed, making faint stars more difficult to observe. Large lenses are also very heavy and deform under their own weight. Some of these problems with refracting telescopes are addressed by avoiding refraction for collecting light and instead using a curved mirror in its place, as devised by Isaac Newton. These telescopes are called reflecting telescopes.

In 1897, the Yerkes Observatory in Wisconsin (USA) built a large refracting telescope with an objective lens that is 40 inches in diameter and has a tube length of 62 feet. (credit: Yerkes Observatory, University of Chicago)

Reflecting Telescopes

Isaac Newton designed the first reflecting telescope around 1670 to solve the problem of chromatic aberration that happens in all refracting telescopes. In chromatic aberration, light of different colors refracts by slightly different amounts in the lens. As a result, a rainbow appears around the image and the image appears blurred. In the reflecting telescope, light rays from a distant source fall upon the surface of a concave mirror fixed at the bottom end of the tube. The use of a mirror instead of a lens eliminates chromatic aberration. The concave mirror focuses the rays on its focal plane. The design problem is how to observe the focused image. Newton used a design in which the focused light from the concave mirror was reflected to one side of the tube into an eyepiece [part (a) of [link]]. This arrangement is common in many amateur telescopes and is called the **Newtonian design**.

Some telescopes reflect the light back toward the middle of the concave mirror using a convex mirror. In this arrangement, the light-gathering concave mirror has a hole in the middle [part (b) of the figure]. The light then is incident on an eyepiece lens. This arrangement of the objective and eyepiece is called the **Cassegrain design**. Most big telescopes, including the Hubble space telescope, are of this design. Other arrangements are also possible. In some telescopes, a light detector is placed right at the spot where light is focused by the curved mirror.



Reflecting telescopes: (a) In the Newtonian design, the eyepiece is located at the side of the telescope; (b) in the Cassegrain design, the eyepiece is located past a hole in the primary mirror.

Most astronomical research telescopes are now of the reflecting type. One of the earliest large telescopes of this kind is the Hale 200-inch (or 5-meter) telescope built on Mount Palomar in southern California, which has a 200 inch-diameter mirror. One of the largest telescopes in the world is the 10-meter Keck telescope at the Keck Observatory on the summit of the dormant Mauna Kea volcano in Hawaii. The Keck Observatory operates two 10-meter telescopes. Each is not a single mirror, but is instead made up of 36 hexagonal mirrors. Furthermore, the two telescopes on the Keck can work together, which increases their power to an effective 85-meter mirror. The Hubble telescope ([link]) is another large reflecting telescope with a 2.4 meter-diameter primary mirror. The Hubble was put into orbit around Earth in 1990.

The Hubble space telescope as seen from the Space Shuttle Discovery. (credit: modification of work by NASA)

The angular magnification *M* of a reflecting telescope is also given by [link]. For a spherical mirror, the focal length is half the radius of curvature, so making a large objective mirror not only helps the telescope collect more light but also increases the magnification of the image.

Summary

- Many optical devices contain more than a single lens or mirror. These are analyzed by considering each element sequentially. The image formed by the first is the object for the second, and so on. The same ray-tracing and thin-lens techniques developed in the previous sections apply to each lens element.
- The overall magnification of a multiple-element system is the product of the linear magnifications of its individual elements times the angular magnification of the eyepiece. For a two-element system with an objective and an eyepiece, this is
  Equation:

  $$M = m^{\text{obj}} M^{\text{eye}}.$$

  where m^{obj} is the linear magnification of the objective and M^{eye} is the angular magnification of the eyepiece.
- The microscope is a multiple-element system that contains more than a single lens or mirror. It allows us to see detail that we could not to see with the unaided eye. Both the eyepiece and objective contribute to the magnification. The magnification of a compound microscope with the image at infinity is
  Equation:

  $$M_{\text{net}} = -\frac{(16\text{ cm})(25\text{ cm})}{f^{\text{obj}} f^{\text{eye}}}.$$

  In this equation, 16 cm is the standardized distance between the image-side focal point of the objective lens and the object-side focal point of the eyepiece, 25 cm is the normal near point distance, f^{obj} and f^{eye} are the focal distances for the objective lens and the eyepiece, respectively.

- Simple telescopes can be made with two lenses. They are used for viewing objects at large distances.
- The angular magnification M for a telescope is given by
  Equation:

  $$M = -\frac{f^{\text{obj}}}{f^{\text{eye}}},$$

  where f^{obj} and f^{eye} are the focal lengths of the objective lens and the eyepiece, respectively.

Key Equations

Image distance in a plane mirror	$d_{\text{o}} = -d_{\text{i}}$
Focal length for a spherical mirror	$f = \frac{R}{2}$
Mirror equation	$\frac{1}{d_{\text{o}}} + \frac{1}{d_{\text{i}}} = \frac{1}{f}$
Magnification of a spherical mirror	$m = \frac{h_{\text{i}}}{h_{\text{o}}} = -\frac{d_{\text{i}}}{d_{\text{o}}}$
Sign convention for mirrors	
Focal length f	$+$ for concave mirror $-$ for convex mirror

Object distance d_o	$+$ for real object $-$ for virtual object
Image distance d_i	$+$ for real image $-$ for virtual image
Magnification m	$+$ for upright image $-$ for inverted image
Apparent depth equation	$h_i = \left(\frac{n_2}{n_1}\right) h_o$
Spherical interface equation	$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{n_2 - n_1}{R}$
The thin-lens equation	$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$
The lens maker’s equation	$\frac{1}{f} = \left(\frac{n_2}{n_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$
The magnification m of an object	$m \equiv \frac{h_i}{h_o} = -\frac{d_i}{d_o}$
Optical power	$P = \frac{1}{f}$
Optical power of thin, closely spaced lenses	$P_{\text{total}} = P_{\text{lens1}} + P_{\text{lens2}} + P_{\text{lens3}} + \cdots$
Angular magnification M of a simple magnifier	$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}}$
Angular magnification of an object a distance L from the eye for a convex lens of focal length	$M = \left(\frac{25\text{ cm}}{L}\right)\left(1 + \frac{L - \ell}{f}\right)$

f held a distance ℓ from the eye	
Range of angular magnification for a given lens for a person with a near point of 25 cm	$\frac{25\text{ cm}}{f} \le M \le 1 + \frac{25\text{ cm}}{f}$
Net magnification of compound microscope	$M_{\text{net}} = m^{\text{obj}} M^{\text{eye}} = -\frac{d_{\text{i}}^{\text{obj}}(f^{\text{eye}}+25\text{ cm})}{f^{\text{obj}} f^{\text{eye}}}$

Conceptual Questions

Exercise:

Problem:

Geometric optics describes the interaction of light with macroscopic objects. Why, then, is it correct to use geometric optics to analyze a microscope’s image?

Solution:

Microscopes create images of macroscopic size, so geometric optics applies.

Exercise:

Problem:

The image produced by the microscope in [link] cannot be projected. Could extra lenses or mirrors project it? Explain.

Exercise:

Problem:

If you want your microscope or telescope to project a real image onto a screen, how would you change the placement of the eyepiece relative to the objective?

Solution:

The eyepiece would be moved slightly farther from the objective so that the image formed by the objective falls just beyond the focal length of the eyepiece.

Problems

Exercise:

Problem:

A microscope with an overall magnification of 800 has an objective that magnifies by 200. (a) What is the angular magnification of the eyepiece? (b) If there are two other objectives that can be used, having magnifications of 100 and 400, what other total magnifications are possible?

Exercise:

Problem:

(a) What magnification is produced by a 0.150 cm-focal length microscope objective that is 0.155 cm from the object being viewed? (b) What is the overall magnification if an $8\ \times$ eyepiece (one that produces an angular magnification of 8.00) is used?

Solution:

a. $\begin{aligned} \frac{1}{d_\text{o}} + \frac{1}{d_\text{i}} &= \frac{1}{f} \Rightarrow d_\text{i} = 4.65\text{ cm} \\ &\Rightarrow m = -30.0 \end{aligned}$;

b. $M_\text{net} = -240$

Exercise:

Problem:

Where does an object need to be placed relative to a microscope for its 0.50 cm-focal length objective to produce a magnification of −400?

Exercise:

Problem:

An amoeba is 0.305 cm away from the 0.300 cm-focal length objective lens of a microscope. (a) Where is the image formed by the objective lens? (b) What is this image’s magnification? (c) An eyepiece with a 2.00-cm focal length is placed 20.0 cm from the objective. Where is the final image? (d) What angular magnification is produced by the eyepiece? (e) What is the overall magnification? (See [link].)

Solution:

a. $\frac{1}{d_\text{o}^\text{obj}} + \frac{1}{d_\text{i}^\text{obj}} = \frac{1}{f^\text{obj}} \Rightarrow d_\text{i}^\text{obj} = 18.3\text{ cm}$ behind the objective lens;
b. $m^\text{obj} = -60.0$;
c. $\begin{aligned} d_\text{o}^\text{eye} &= 1.70\text{ cm} \\ d_\text{i}^\text{eye} &= -11.3\text{ cm} \end{aligned}$
in front of the eyepiece; d. $M^\text{eye} = 13.5$;
e. $M_\text{net} = -810$

Exercise:

Problem:

Unreasonable Results Your friends show you an image through a microscope. They tell you that the microscope has an objective with a 0.500-cm focal length and an eyepiece with a 5.00-cm focal length. The resulting overall magnification is 250,000. Are these viable values for a microscope?

Unless otherwise stated, the lens-to-retina distance is 2.00 cm.

Exercise:

Problem:

What is the angular magnification of a telescope that has a 100 cm-focal length objective and a 2.50 cm-focal length eyepiece?

Solution:

$M = -40.0$

Exercise:

Problem:

Find the distance between the objective and eyepiece lenses in the telescope in the above problem needed to produce a final image very far from the observer, where vision is most relaxed. Note that a telescope is normally used to view very distant objects.

Exercise:

Problem:

A large reflecting telescope has an objective mirror with a 10.0-m radius of curvature. What angular magnification does it produce when a 3.00 m-focal length eyepiece is used?

Solution:

$f^{\text{obj}} = \frac{R}{2}, M = -1.67$

Exercise:

Problem:

A small telescope has a concave mirror with a 2.00-m radius of curvature for its objective. Its eyepiece is a 4.00 cm-focal length lens. (a) What is the telescope’s angular magnification? (b) What angle is subtended by a 25,000 km-diameter sunspot? (c) What is the angle of its telescopic image?

Exercise:

Problem:

A $7.5\ \times$ binocular produces an angular magnification of −7.50, acting like a telescope. (Mirrors are used to make the image upright.) If the binoculars have objective lenses with a 75.0-cm focal length, what is the focal length of the eyepiece lenses?

Solution:

$$M=-\frac{f^{\text{obj}}}{f^{\text{eye}}},f^{\text{eye}}=+10.0\ \text{cm}$$

Exercise:

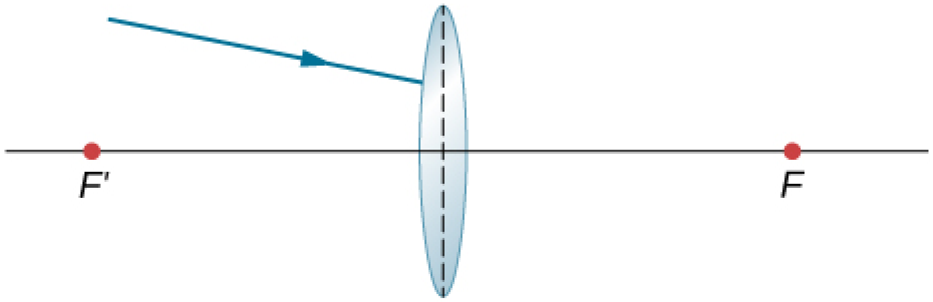
Problem:

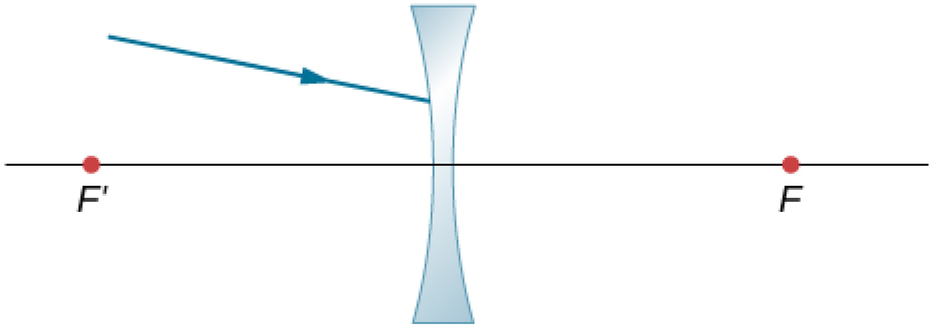
Construct Your Own Problem Consider a telescope of the type used by Galileo, having a convex objective and a concave eyepiece as illustrated in part (a) of [link]. Construct a problem in which you calculate the location and size of the image produced. Among the things to be considered are the focal lengths of the lenses and their relative placements as well as the size and location of the object. Verify that the angular magnification is greater than one. That is, the angle subtended at the eye by the image is greater than the angle subtended by the object.

Exercise:

Problem:

Trace rays to find which way the given ray will emerge after refraction through the thin lens in the following figure. Assume thin-lens approximation. (*Hint*: Pick a point *P* on the given ray in each case. Treat that point as an object. Now, find its image *Q*. Use the rule: All rays on the other side of the lens will either go through *Q* or appear to be coming from *Q*.)





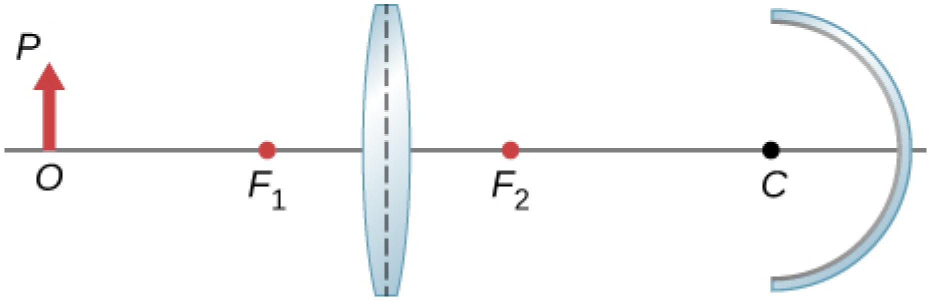
Solution:

Answers will vary.

Exercise:

Problem:

Copy and draw rays to find the final image in the following diagram. (*Hint*: Find the intermediate image through lens alone. Use the intermediate image as the object for the mirror and work with the mirror alone to find the final image.)



Exercise:

Problem:

A concave mirror of radius of curvature 10 cm is placed 30 cm from a thin convex lens of focal length 15 cm. Find the location and magnification of a small bulb sitting 50 cm from the lens by using the algebraic method.

Solution:

12 cm to the left of the mirror, $m = 3/5$

Exercise:

Problem:

An object of height 3 cm is placed at 25 cm in front of a converging lens of focal length 20 cm. Behind the lens there is a concave mirror of focal length 20 cm. The distance between the lens and the mirror is 5 cm. Find the location, orientation and size of the final image.

Exercise:

Problem:

An object of height 3 cm is placed at a distance of 25 cm in front of a converging lens of focal length 20 cm, to be referred to as the first lens. Behind the lens there is another converging lens of focal length 20 cm placed 10 cm from the first lens. There is a concave mirror of focal length 15 cm placed 50 cm from the second lens. Find the location, orientation, and size of the final image.

Solution:

27 cm in front of the mirror, $m = 0.6,\ h_i = 1.76$ cm, orientation upright

Exercise:

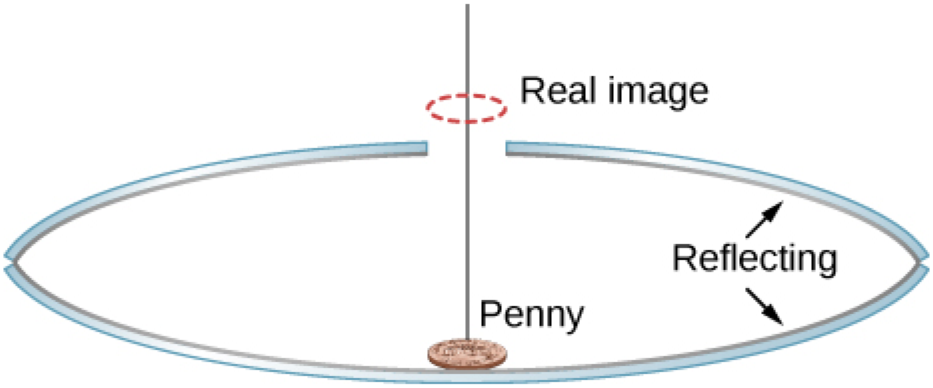
Problem:

An object of height 2 cm is placed at 50 cm in front of a diverging lens of focal length 40 cm. Behind the lens, there is a convex mirror of focal length 15 cm placed 30 cm from the converging lens. Find the location, orientation, and size of the final image.

Exercise:

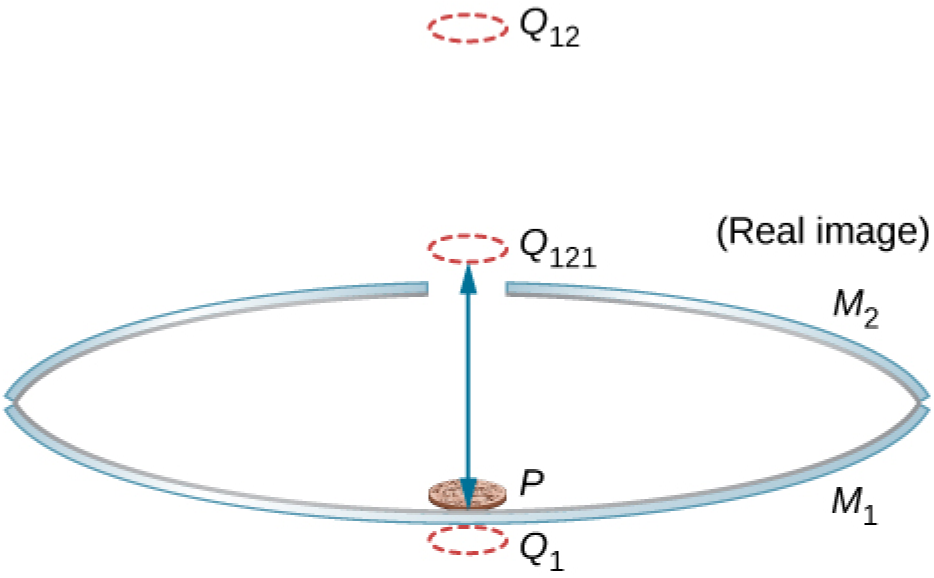
Problem:

Two concave mirrors are placed facing each other. One of them has a small hole in the middle. A penny is placed on the bottom mirror (see the following figure). When you look from the side, a real image of the penny is observed above the hole. Explain how that could happen.



Solution:

The following figure shows three successive images beginning with the image Q_1 in mirror M_1. Q_1 is the image in mirror M_1, whose image in mirror M_2 is Q_{12} whose image in mirror M_1 is the real image Q_{121}.



Exercise:

Problem:

A lamp of height 5 cm is placed 40 cm in front of a converging lens of focal length 20 cm. There is a plane mirror 15 cm behind the lens. Where would you find the image when you look in the mirror?

Exercise:

Problem:

Parallel rays from a faraway source strike a converging lens of focal length 20 cm at an angle of 15 degrees with the horizontal direction. Find the vertical position of the real image observed on a screen in the focal plane.

Solution:

5.4 cm from the axis

Exercise:

Problem:

Parallel rays from a faraway source strike a diverging lens of focal length 20 cm at an angle of 10 degrees with the horizontal direction. As you look through the lens, where in the vertical plane the image would appear?

Exercise:

Problem:

A light bulb is placed 10 cm from a plane mirror, which faces a convex mirror of radius of curvature 8 cm. The plane mirror is located at a distance of 30 cm from the vertex of the convex mirror. Find the location of two images in the convex mirror. Are there other images? If so, where are they located?

Solution:

Let the vertex of the concave mirror be the origin of the coordinate system. Image 1 is at −10/3 cm (−3.3 cm), image 2 is at −40/11 cm (−3.6 cm). These serve as objects for subsequent images, which are at −310/83 cm (−3.7 cm), −9340/2501 cm (−3.7 cm), −140,720/37,681 cm (−3.7 cm). All remaining images are at approximately −3.7 cm.

Exercise:

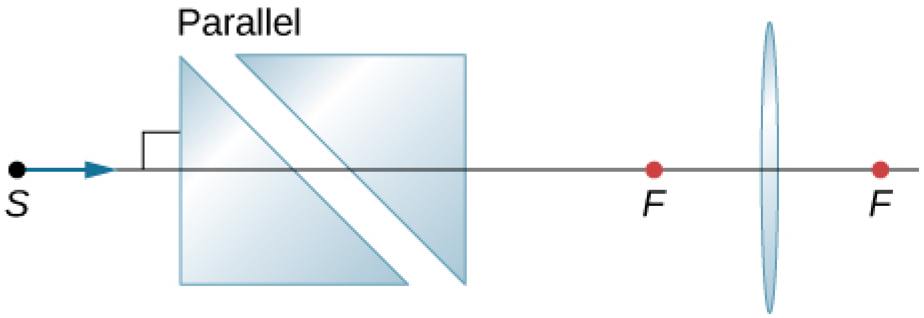
Problem:

A point source of light is 50 cm in front of a converging lens of focal length 30 cm. A concave mirror with a focal length of 20 cm is placed 25 cm behind the lens. Where does the final image form, and what are its orientation and magnification?

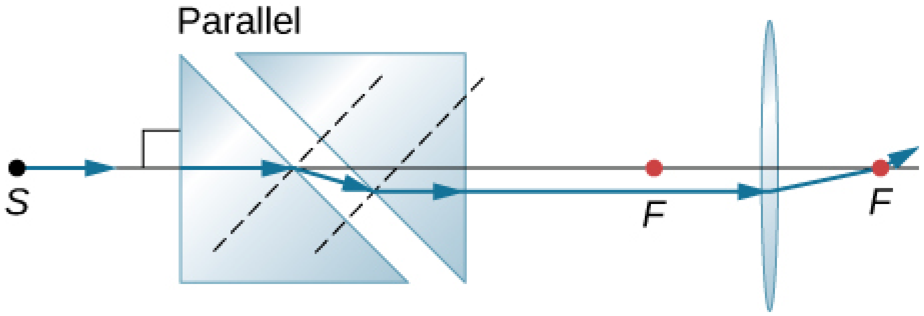
Exercise:

Problem:

Copy and trace to find how a horizontal ray from *S* comes out after the lens. Use $n_{\text{glass}} = 1.5$ for the prism material.



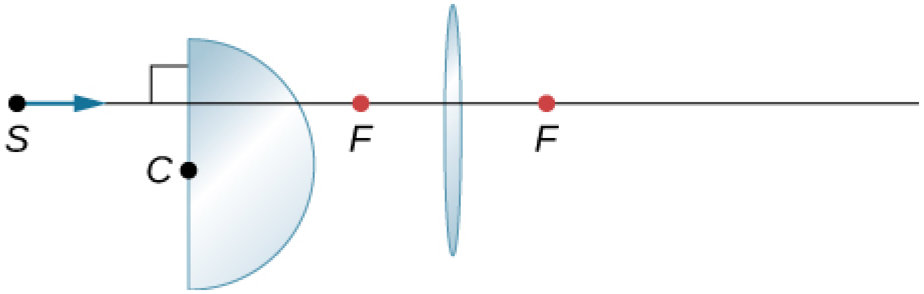
Solution:



Exercise:

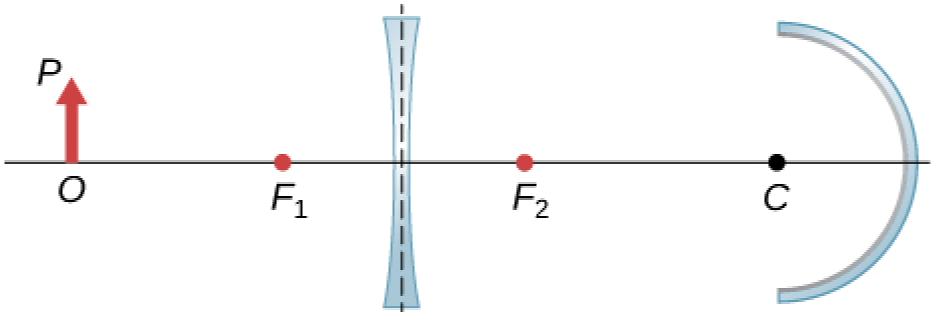
Problem:

Copy and trace how a horizontal ray from S comes out after the lens. Use $n = 1.55$ for the glass.

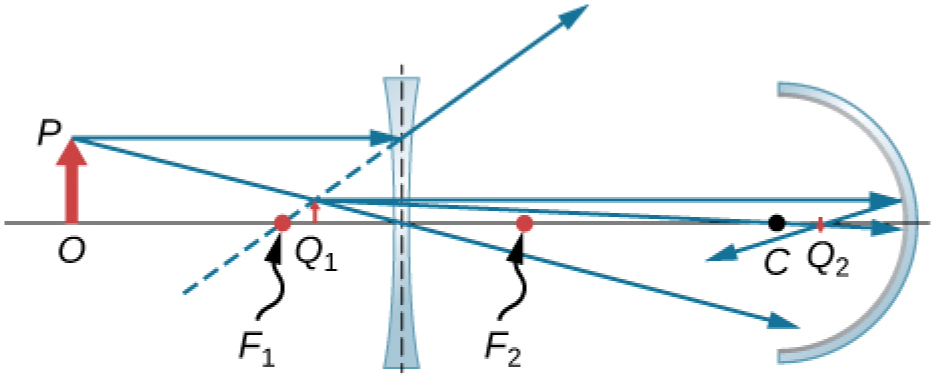


Exercise:

Problem: Copy and draw rays to figure out the final image.



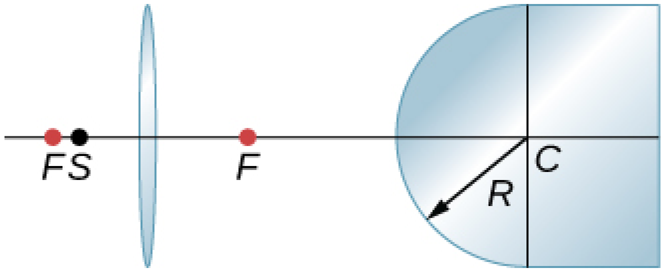
Solution:



Exercise:

Problem:

By ray tracing or by calculation, find the place inside the glass where rays from *S* converge as a result of refraction through the lens and the convex air-glass interface. Use a ruler to estimate the radius of curvature.



Exercise:

Problem:

A diverging lens has a focal length of 20 cm. What is the power of the lens in diopters?

Solution:

−5 D

Exercise:

Problem:

Two lenses of focal lengths of f_1 and f_2 are glued together with transparent material of negligible thickness. Show that the total power of the two lenses simply add.

Exercise:

Problem:

What will be the angular magnification of a convex lens with the focal length 2.5 cm?

Solution:

11

Exercise:

Problem:

What will be the formula for the angular magnification of a convex lens of focal length f if the eye is very close to the lens and the near point is located a distance D from the eye?

Additional Problems

Exercise:

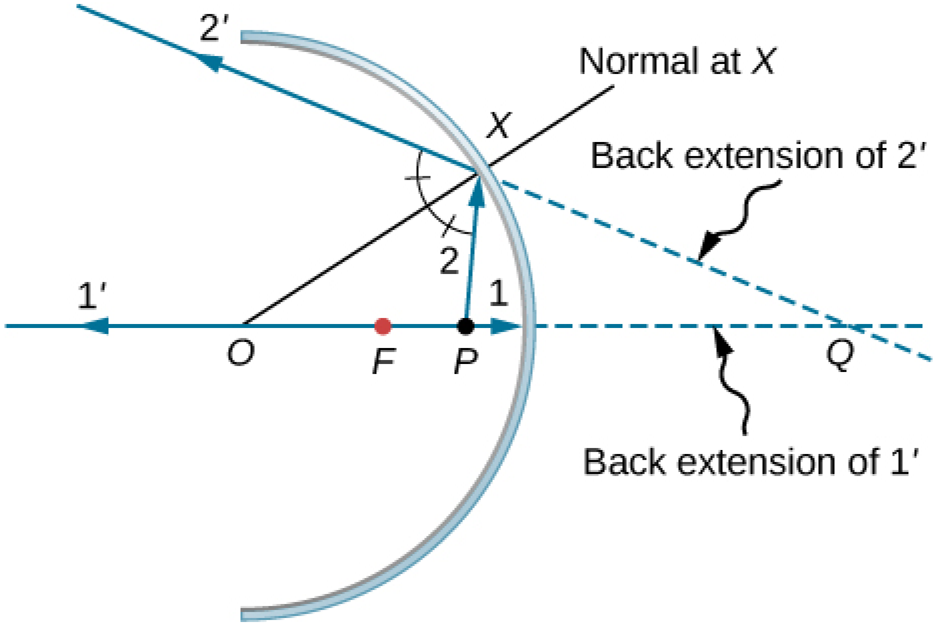
Problem:

Use a ruler and a protractor to draw rays to find images in the following cases.

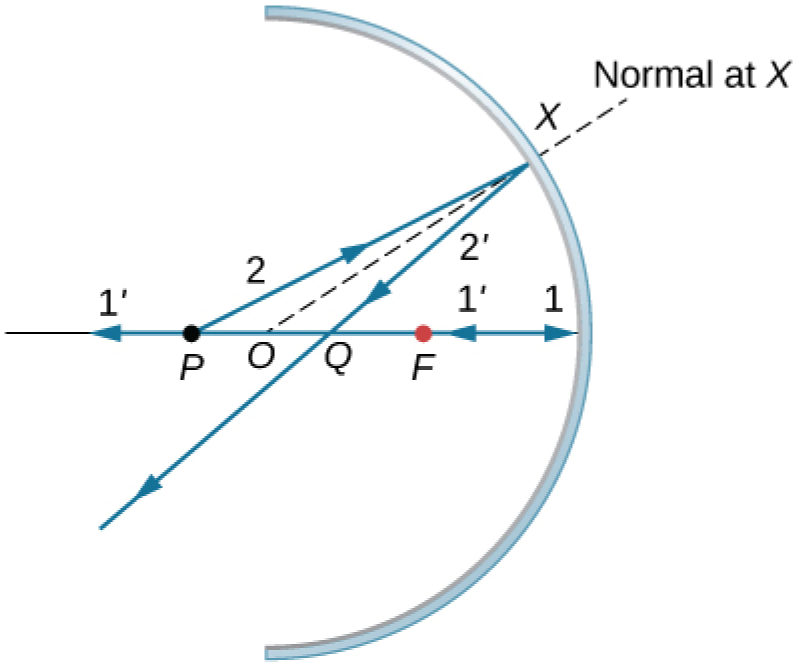
(a) A point object located on the axis of a concave mirror located at a point within the focal length from the vertex.
(b) A point object located on the axis of a concave mirror located at a point farther than the focal length from the vertex.
(c) A point object located on the axis of a convex mirror located at a point within the focal length from the vertex.
(d) A point object located on the axis of a convex mirror located at a point farther than the focal length from the vertex.
(e) Repeat (a)–(d) for a point object off the axis.

Solution:

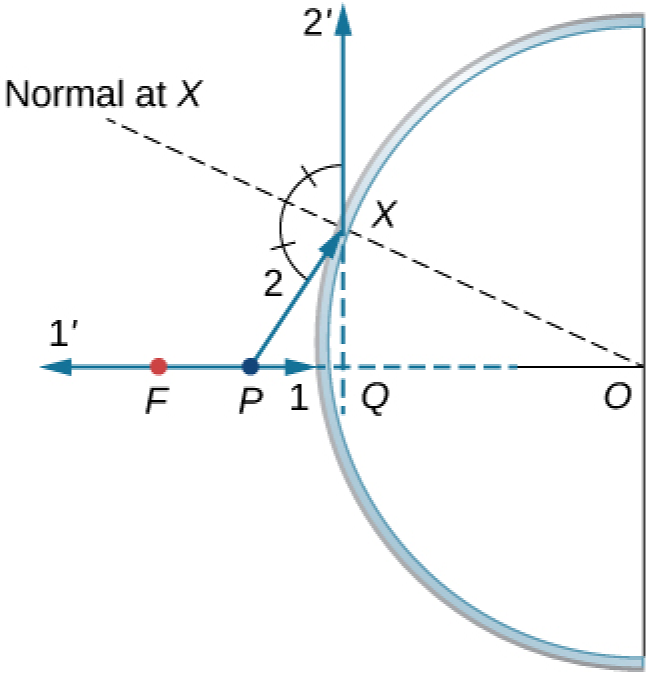
a.



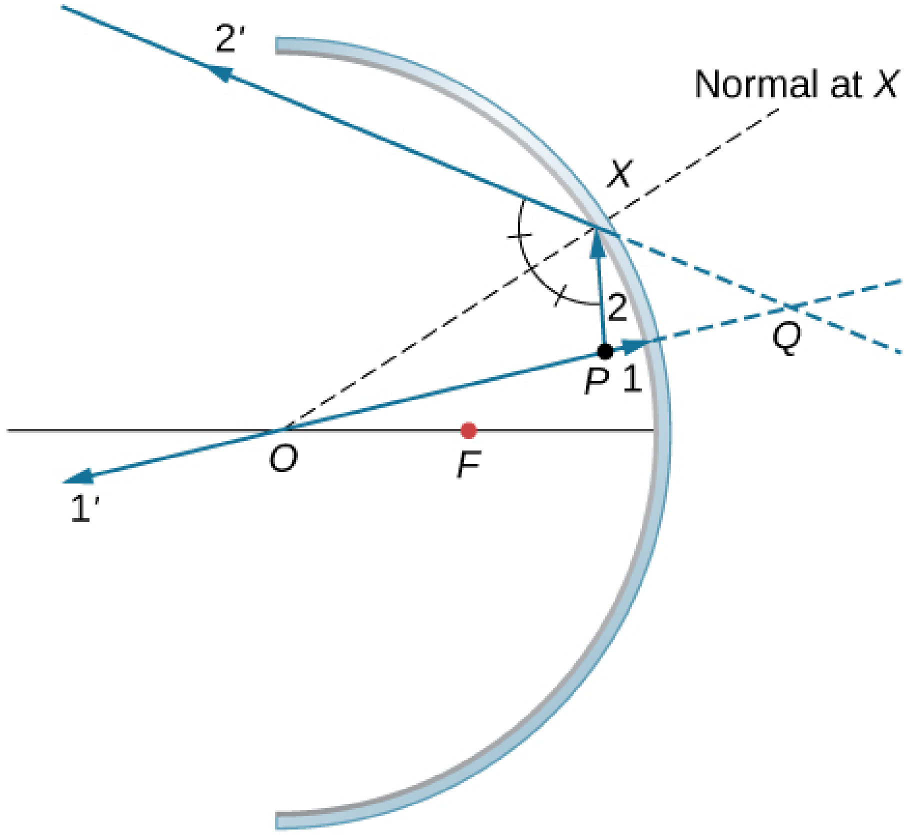
b.

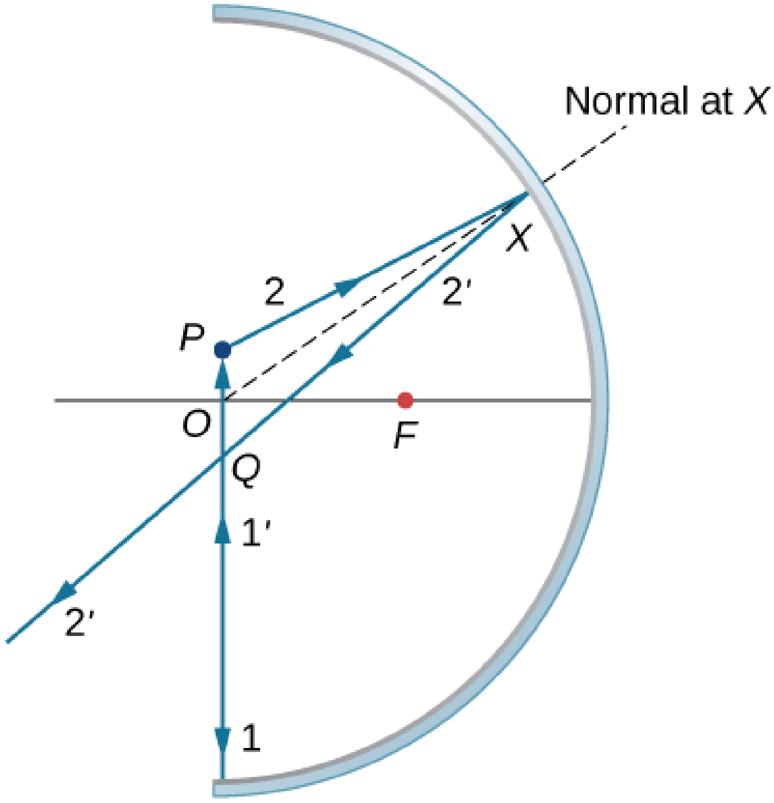


c.



d. similar to the previous picture but with point *P* outside the focal length; e. Repeat (a)–(d) for a point object off the axis. For a point object placed off axis in front of a concave mirror corresponding to parts (a) and (b), the case for convex mirror left as exercises.


2′
Normal at *X*
X
2
Q
P
1
O
F
1′


Normal at *X*
X
2
2′
P
O
F
Q
1′
2′
1

Exercise:

Problem:

Where should a 3 cm tall object be placed in front of a concave mirror of radius 20 cm so that its image is real and 2 cm tall?

Exercise:

Problem:

A 3 cm tall object is placed 5 cm in front of a convex mirror of radius of curvature 20 cm. Where is the image formed? How tall is the image? What is the orientation of the image?

Solution:

$d_{\text{i}} = -10/3\ \text{cm}, h_{\text{i}} = 2\ \text{cm}$, upright

Exercise:

Problem:

You are looking for a mirror so that you can see a four-fold magnified virtual image of an object when the object is placed 5 cm from the vertex of the mirror. What kind of mirror you will need? What should be the radius of curvature of the mirror?

Exercise:

Problem: Derive the following equation for a convex mirror:

$$\frac{1}{VO} - \frac{1}{VI} = -\frac{1}{VF},$$

where *VO* is the distance to the object *O* from vertex *V*, *VI* the distance to the image *I* from *V*, and *VF* is the distance to the focal point *F* from *V*. (*Hint*: use two sets of similar triangles.)

Solution:

proof

Exercise:

Problem:

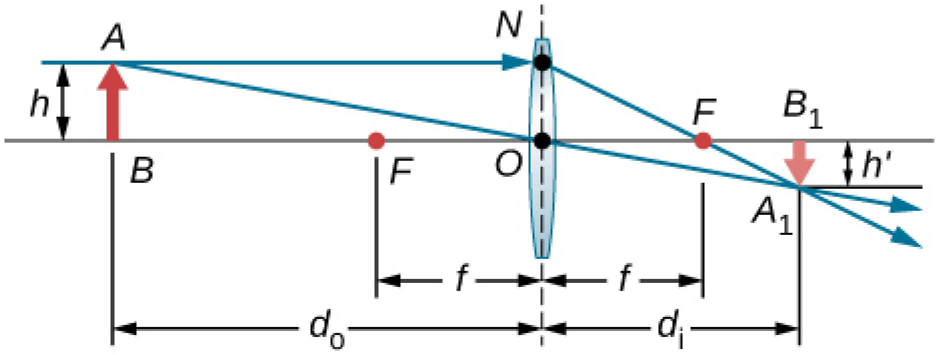
(a) Draw rays to form the image of a vertical object on the optical axis and farther than the focal point from a converging lens. (b) Use plane geometry in your figure and prove that the magnification *m* is given by $m=\frac{h_{\text{i}}}{h_{\text{o}}}=-\frac{d_{\text{i}}}{d_{\text{o}}}$.

Exercise:

Problem:

Use another ray-tracing diagram for the same situation as given in the previous problem to derive the thin-lens equation, $\frac{1}{d_{\text{o}}}+\frac{1}{d_{\text{i}}}=\frac{1}{f}$.

Solution:



Triangles *BAO* and B_1A_1O are similar triangles. Thus, $\frac{A_1B_1}{AB}=\frac{d_{\text{i}}}{d_{\text{o}}}$ .Triangles *NOF* and B_1A_1F are similar triangles. Thus, $\frac{NO}{f}=\frac{A_1B_1}{d_{\text{i}}-f}$. Noting that $NO=AB$ gives $\frac{AB}{f}=\frac{A_1B_1}{d_{\text{i}}-f}$ or $\frac{AB}{A_1B_1}=\frac{f}{d_{\text{i}}-f}$. Inverting this gives $\frac{A_1B_1}{AB}=\frac{d_{\text{i}}-f}{f}$. Equating the two expressions for the ratio $\frac{A_1B_1}{AB}$ gives $\frac{d_{\text{i}}}{d_{\text{o}}}=\frac{d_{\text{i}}-f}{f}$. Dividing through by d_{i} gives $\frac{1}{d_{\text{o}}}=\frac{1}{f}-\frac{1}{d_{\text{i}}}$ or $\frac{1}{d_{\text{o}}}+\frac{1}{d_{\text{i}}}=\frac{1}{f}$.

Exercise:

Problem:

You photograph a 2.0-m-tall person with a camera that has a 5.0 cm-focal length lens. The image on the film must be no more than 2.0 cm high. (a) What is the closest distance the person can stand to the lens? (b) For this distance, what should be the distance from the lens to the film?

Exercise:

Problem:

Find the focal length of a thin plano-convex lens. The front surface of this lens is flat, and the rear surface has a radius of curvature of $R_2 = -35\text{ cm}$. Assume that the index of refraction of the lens is 1.5.

Solution:

70 cm

Exercise:

Problem:

Find the focal length of a meniscus lens with $R_1 = 20\text{ cm}$ and $R_2 = 15\text{ cm}$. Assume that the index of refraction of the lens is 1.5.

Exercise:

Problem:

A nearsighted man cannot see objects clearly beyond 20 cm from his eyes. How close must he stand to a mirror in order to see what he is doing when he shaves?

Solution:

The plane mirror has an infinite focal point, so that $d_\text{i} = -d_\text{o}$. The total apparent distance of the man in the mirror will be his actual distance, plus the apparent image distance, or $d_\text{o} + (-d_\text{i}) = 2d_\text{o}$. If this distance must be less than 20 cm, he should stand at $d_\text{o} = 10\text{ cm}$.

Exercise:

Problem:

A mother sees that her child's contact lens prescription is 0.750 D. What is the child's near point?

Exercise:

Problem:

Repeat the previous problem for glasses that are 2.20 cm from the eyes.

Solution:

Here we want $d_{\text{o}} = 25\text{ cm} - 2.20\text{ cm} = 0.228\text{ m}$. If $x = \text{near point}$, $d_{\text{i}} = -(x - 0.0220\text{ m})$. Thus, $P = \frac{1}{d_{\text{o}}} + \frac{1}{d_{\text{i}}} = \frac{1}{0.228\text{ m}} + \frac{1}{x - 0.0220\text{ m}}$. Using $P = 0.75\text{ D}$ gives $x = 0.253\text{ m}$, so the near point is 25.3 cm.

Exercise:

Problem:

The contact-lens prescription for a nearsighted person is −4.00 D and the person has a far point of 22.5 cm. What is the power of the tear layer between the cornea and the lens if the correction is ideal, taking the tear layer into account?

Exercise:

Problem:

Unreasonable Results A boy has a near point of 50 cm and a far point of 500 cm. Will a −4.00 D lens correct his far point to infinity?

Solution:

Assuming a lens at 2.00 cm from the boy's eye, the image distance must be $d_{\text{i}} = -(500\text{ cm} - 2.00\text{ cm}) = -498\text{ cm}$. For an infinite-distance object, the required power is $P = \frac{1}{d_\text{i}} = -0.200\,\text{D}$. Therefore, the $-4.00\,\text{D}$ lens will correct the nearsightedness.

Exercise:

Problem:

Find the angular magnification of an image by a magnifying glass of $f = 5.0\,\text{cm}$ if the object is placed $d_\text{o} = 4.0\,\text{cm}$ from the lens and the lens is close to the eye.

Exercise:

Problem:

Let objective and eyepiece of a compound microscope have focal lengths of 2.5 cm and 10 cm, respectively and be separated by 12 cm. A $70\text{-}\mu\text{m}$ object is placed 6.0 cm from the objective. How large is the virtual image formed by the objective-eyepiece system?

Solution:

$87\,\mu\text{m}$

Exercise:

Problem:

Draw rays to scale to locate the image at the retina if the eye lens has a focal length 2.5 cm and the near point is 24 cm. (*Hint*: Place an object at the near point.)

Exercise:

Problem:

The objective and the eyepiece of a microscope have the focal lengths 3 cm and 10 cm respectively. Decide about the distance between the objective and the eyepiece if we need a $10\,\times$ magnification from the objective/eyepiece compound system.

Solution:

Use, $M_{\text{net}} = -\frac{d_{\text{i}}^{\text{obj}}(f^{\text{eye}}+25\text{ cm})}{f^{\text{obj}}f^{\text{eye}}}$. The image distance for the objective is $d_{\text{i}}^{\text{obj}} = -\frac{M_{\text{net}}f^{\text{obj}}f^{\text{eye}}}{f^{\text{eye}}+25\text{ cm}}$. Using $f^{\text{obj}} = 3.0\text{ cm}, f^{\text{eye}} = 10\text{ cm},\text{and }M = -10\text{ gives }d_{\text{i}}^{\text{obj}} = 8.6\text{ cm}.$ We want this image to be at the focal point of the eyepiece so that the eyepiece forms an image at infinity for comfortable viewing. Thus, the distance *d* between the lenses should be $d = f^{\text{eye}} + d_{\text{i}}^{\text{obj}} = 10\text{ cm} + 8.6\text{ cm} = 19\text{ cm}.$

Exercise:

Problem:

A far-sighted person has a near point of 100 cm. How far in front or behind the retina does the image of an object placed 25 cm from the eye form? Use the cornea to retina distance of 2.5 cm.

Exercise:

Problem:

A near-sighted person has afar point of 80 cm. (a) What kind of corrective lens the person will need if the lens is to be placed 1.5 cm from the eye? (b) What would be the power of the contact lens needed? Assume distance to contact lens from the eye to be zero.

Solution:

a. focal length of the corrective lens $f_{\text{c}} = -80\text{ cm}$; b. −1.25 D

Exercise:

Problem:

In a reflecting telescope the objective is a concave mirror of radius of curvature 2 m and an eyepiece is a convex lens of focal length 5 cm. Find the apparent size of a 25-m tree at a distance of 10 km that you would perceive when looking through the telescope.

Exercise:

Problem:

Two stars that are $10^9\mathrm{km}$ apart are viewed by a telescope and found to be separated by an angle of 10^{-5} $\mathrm{radians}$. If the eyepiece of the telescope has a focal length of 1.5 cm and the objective has a focal length of 3 meters, how far away are the stars from the observer?

Solution:

$2\ \times\ 10^{16}\ \mathrm{km}$

Exercise:

Problem:

What is the angular size of the Moon if viewed from a binocular that has a focal length of 1.2 cm for the eyepiece and a focal length of 8 cm for the objective? Use the radius of the moon $1.74\ \times\ 10^6\mathrm{m}$ and the distance of the moon from the observer to be $3.8\ \times\ 10^8\mathrm{m}$.

Exercise:

Problem:

An unknown planet at a distance of $10^{12}\mathrm{m}$ from Earth is observed by a telescope that has a focal length of the eyepiece of 1 cm and a focal length of the objective of 1 m. If the far away planet is seen to subtend an angle of 10^{-5} radian at the eyepiece, what is the size of the planet?

Solution:

$10^5\ \mathrm{m}$

Glossary

Cassegrain design
: arrangement of an objective and eyepiece such that the light-gathering concave mirror has a hole in the middle, and light then is incident on

an eyepiece lens

compound microscope
microscope constructed from two convex lenses, the first serving as the eyepiece and the second serving as the objective lens

eyepiece
lens or combination of lenses in an optical instrument nearest to the eye of the observer

net magnification
(M_{net}) of the compound microscope is the product of the linear magnification of the objective and the angular magnification of the eyepiece

Newtonian design
arrangement of an objective and eyepiece such that the focused light from the concave mirror was reflected to one side of the tube into an eyepiece

objective
lens nearest to the object being examined.

Traveling Waves

By the end of this section, you will be able to:

- Describe the basic characteristics of wave motion
- Define the terms wavelength, amplitude, period, frequency, and wave speed
- Explain the difference between longitudinal and transverse waves, and give examples of each type
- List the different types of waves

We saw in Oscillations that oscillatory motion is an important type of behavior that can be used to model a wide range of physical phenomena. Oscillatory motion is also important because oscillations can generate waves, which are of fundamental importance in physics. Many of the terms and equations we studied in the chapter on oscillations apply equally well to wave motion ([link]).

An ocean wave is probably the first picture that comes to mind when you hear the word “wave.” Although this breaking wave, and ocean waves in general, have apparent similarities to the basic wave characteristics we will discuss, the mechanisms driving ocean waves are highly complex and beyond the scope of this chapter. It may seem natural, and even advantageous, to apply the concepts in this chapter to ocean waves, but ocean waves are nonlinear, and the simple models presented in this chapter do not fully explain them. (credit: Steve Jurvetson)

Types of Waves

A **wave** is a disturbance that propagates, or moves from the place it was created. There are three basic types of waves: mechanical waves, electromagnetic waves, and matter waves.

Basic **mechanical waves** are governed by Newton’s laws and require a medium. A medium is the substance a mechanical waves propagates through, and the medium produces an elastic restoring force when it is deformed. Mechanical waves transfer energy and momentum, without transferring mass. Some examples of mechanical waves are water waves, sound waves, and seismic waves. The medium for water waves is water; for sound waves, the medium is usually air. (Sound waves can travel in other media as well; we will look at that in more detail in Sound.) For surface water waves, the disturbance occurs on the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker or a vibrating tuning fork. In both cases, the disturbance is the oscillation of the molecules of the fluid. In mechanical waves, energy and momentum transfer with the motion of the wave, whereas the mass oscillates around an equilibrium point. (We discuss this in Energy and Power of a Wave.) Earthquakes generate seismic waves from several types of disturbances, including the disturbance of Earth’s surface and pressure disturbances under the surface. Seismic waves travel through the solids and liquids that form Earth. In this chapter, we focus on mechanical waves.

Electromagnetic waves are associated with oscillations in electric and magnetic fields and do not require a medium. Examples include gamma rays, X-rays, ultraviolet waves, visible light, infrared waves, microwaves, and radio waves. Electromagnetic waves can travel through a vacuum at the speed of light, $v = c = 2.99792458 \times 10^8 \text{ m/s}$. For example, light from distant stars travels through the vacuum of space and reaches Earth. Electromagnetic waves have some characteristics that are similar to mechanical waves; they are covered in more detail in Electromagnetic Waves.

Matter waves are a central part of the branch of physics known as quantum mechanics. These waves are associated with protons, electrons, neutrons, and other fundamental particles found in nature. The theory that all types of matter have wave-like properties was first proposed by Louis de Broglie in 1924. Matter waves are discussed in Photons and Matter Waves.

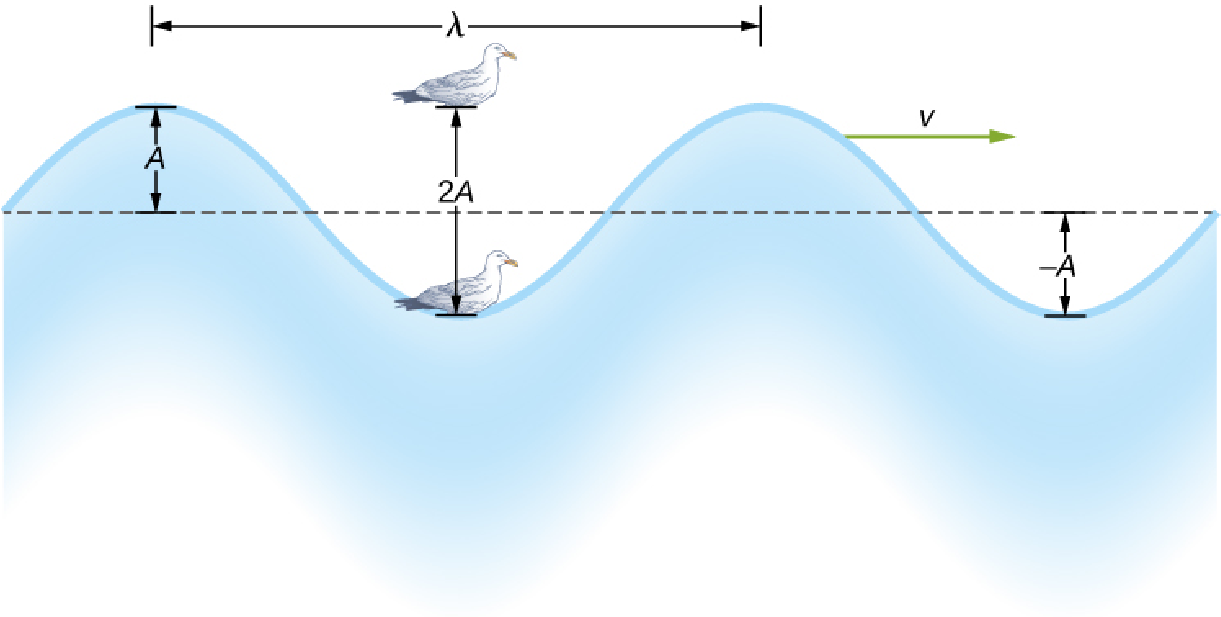
Mechanical Waves

Mechanical waves exhibit characteristics common to all waves, such as amplitude, wavelength, period, frequency, and energy. All wave characteristics can be described by a small set of underlying principles.

The simplest mechanical waves repeat themselves for several cycles and are associated with simple harmonic motion. These simple harmonic waves can be modeled using some combination of sine and cosine functions. For example, consider the simplified surface water wave that moves across the surface of water as illustrated in [link]. Unlike complex ocean waves, in surface water waves, the medium, in this case water, moves vertically, oscillating up and down, whereas the disturbance of the wave moves horizontally through the medium. In [link], the waves causes a seagull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The crest is the highest point of the wave, and the trough is the lowest part of the wave. The time for one complete oscillation of the up-and-down motion is the wave’s period T. The wave’s frequency is the number of waves that pass through a point per unit time and is equal to $f = 1/T$. The period can be expressed using any convenient unit of time but is usually measured in seconds; frequency is usually measured in hertz (Hz), where $1\ \mathrm{Hz} = 1\ \mathrm{s}^{-1}$.

The length of the wave is called the **wavelength** and is represented by the Greek letter lambda (λ), which is measured in any convenient unit of length, such as a centimeter or meter. The wavelength can be measured between any two similar points along the medium that have the same height and the same slope. In [link], the wavelength is shown measured between two crests. As stated above, the period of the wave is equal to the time for one oscillation, but it is also equal to the time for one wavelength to pass through a point along the wave's path.

The amplitude of the wave (A) is a measure of the maximum displacement of the medium from its equilibrium position. In the figure, the equilibrium position is indicated by the dotted line, which is the height of the water if there were no waves moving through it. In this case, the wave is symmetrical, the crest of the wave is a distance $+A$ above the equilibrium position, and the trough is a distance $-A$ below the equilibrium position. The units for the amplitude can be centimeters or meters, or any convenient unit of distance.



An idealized surface water wave passes under a seagull that bobs up and down in simple harmonic motion. The wave has a wavelength λ, which is the distance between adjacent identical parts of the wave. The amplitude A of the wave is the maximum displacement of the wave from the equilibrium position, which is indicated by the dotted line. In this example, the medium moves up and down, whereas the disturbance of the surface propagates parallel to the surface at a speed v.

The water wave in the figure moves through the medium with a propagation velocity $\vec{\mathbf{v}}$. The magnitude of the **wave velocity** is the distance the wave travels in a given time, which is one wavelength in the time of one period, and the **wave speed** is the magnitude of wave velocity. In equation form, this is

Note:
Equation:

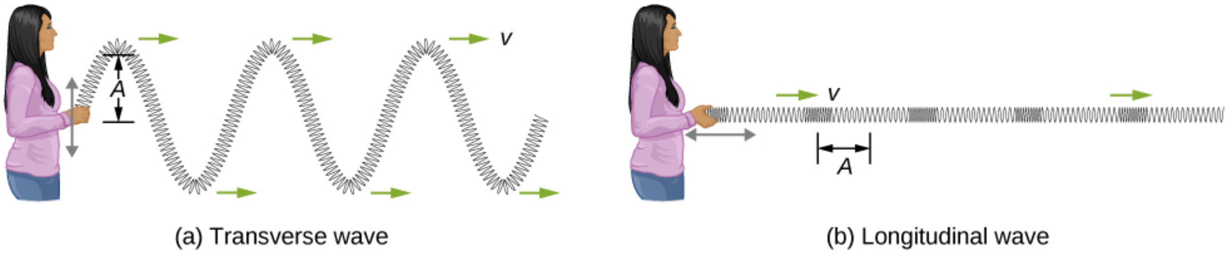
$$v = \frac{\lambda}{T} = \lambda f.$$

This fundamental relationship holds for all types of waves. For water waves, v is the speed of a surface wave; for sound, v is the speed of sound; and for visible light, v is the speed of light.

Transverse and Longitudinal Waves

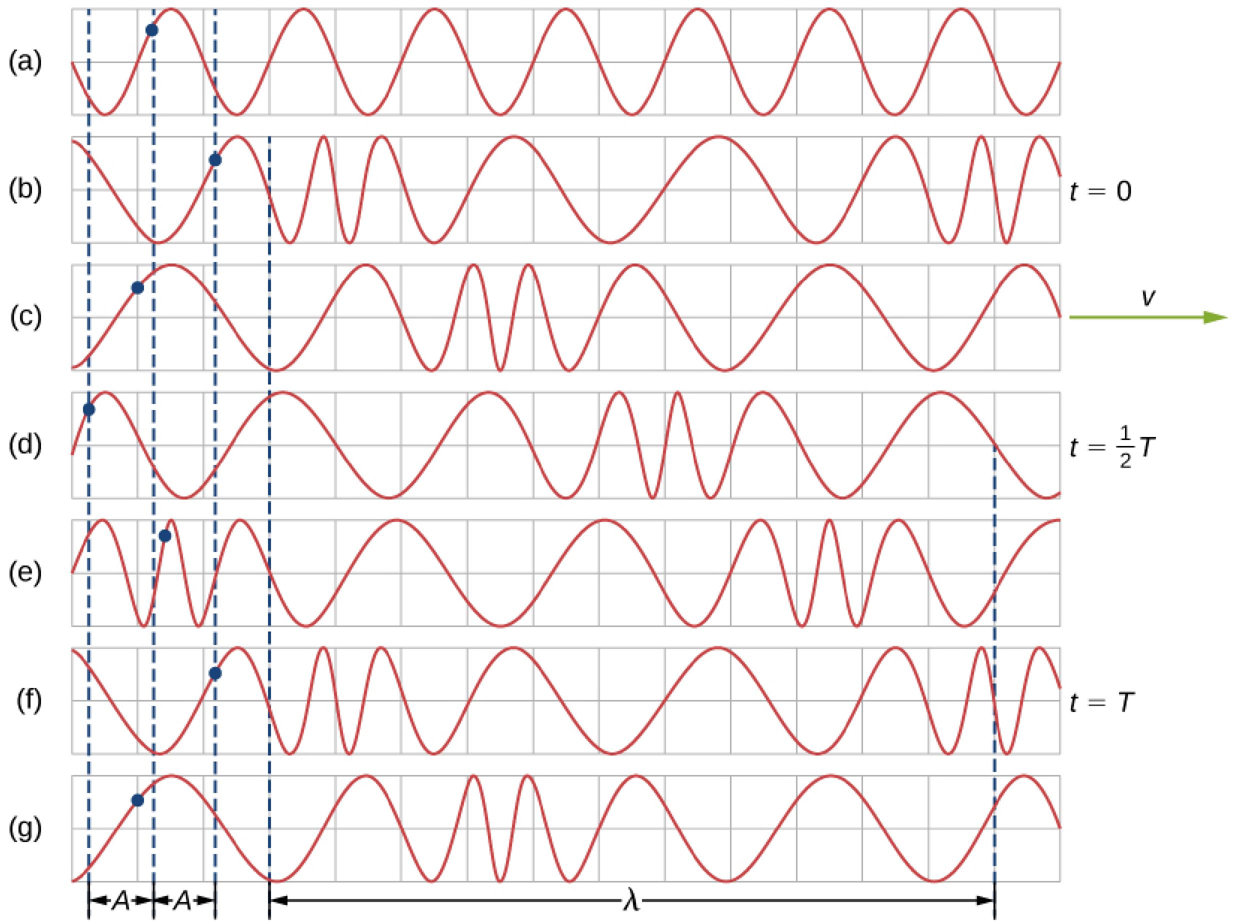
We have seen that a simple mechanical wave consists of a periodic disturbance that propagates from one place to another through a medium. In [link](a), the wave propagates in the horizontal direction, whereas the medium is disturbed in the vertical direction. Such a wave is called a

transverse wave. In a transverse wave, the wave may propagate in any direction, but the disturbance of the medium is perpendicular to the direction of propagation. In contrast, in a **longitudinal wave** or compressional wave, the disturbance is parallel to the direction of propagation. [link](b) shows an example of a longitudinal wave. The size of the disturbance is its amplitude *A* and is completely independent of the speed of propagation *v*.



(a) In a transverse wave, the medium oscillates perpendicular to the wave velocity. Here, the spring moves vertically up and down, while the wave propagates horizontally to the right. (b) In a longitudinal wave, the medium oscillates parallel to the propagation of the wave. In this case, the spring oscillates back and forth, while the wave propagates to the right.

A simple graphical representation of a section of the spring shown in [link](b) is shown in [link]. [link](a) shows the equilibrium position of the spring before any waves move down it. A point on the spring is marked with a blue dot. [link](b) through (g) show snapshots of the spring taken one-quarter of a period apart, sometime after the end of` the spring is oscillated back and forth in the *x*-direction at a constant frequency. The disturbance of the wave is seen as the compressions and the expansions of the spring. Note that the blue dot oscillates around its equilibrium position a distance *A*, as the longitudinal wave moves in the positive *x*-direction with a constant speed. The distance *A* is the amplitude of the wave. The *y*-position of the dot does not change as the wave moves through the spring. The wavelength of the wave is measured in part (d). The wavelength depends on the speed of the wave and the frequency of the driving force.



(a) This is a simple, graphical representation of a section of the stretched spring shown in [link](b), representing the spring’s equilibrium position before any waves are induced on the spring. A point on the spring is marked by a blue dot. (b–g) Longitudinal waves are created by oscillating the end of the spring (not shown) back and forth along the *x*-axis. The longitudinal wave, with a wavelength λ, moves along the spring in the +*x*-direction with a wave speed *v*. For convenience, the wavelength is measured in (d). Note that the point on the spring that was marked with the blue dot moves back and forth a distance *A* from the equilibrium position, oscillating around the equilibrium position of the point.

Waves may be transverse, longitudinal, or a combination of the two. Examples of transverse waves are the waves on stringed instruments or surface waves on water, such as ripples moving on a pond. Sound waves in air and water are longitudinal. With sound waves, the disturbances are periodic variations in pressure that are transmitted in fluids. Fluids do not have appreciable shear strength, and for this reason, the sound waves in them are longitudinal waves. Sound in solids can have both longitudinal and transverse components, such as those in a seismic wave. Earthquakes generate seismic waves under Earth’s surface with both longitudinal and transverse components (called compressional or P-waves and shear or S-waves, respectively). The components of seismic waves have important individual characteristics—they propagate at different speeds, for example. Earthquakes also have surface waves that are similar to surface waves on water. Ocean waves also have both transverse and longitudinal components.

Example:

Wave on a String

A student takes a 30.00-m-long string and attaches one end to the wall in the physics lab. The student then holds the free end of the rope, keeping the tension constant in the rope. The student then begins to send waves down the string by moving the end of the string up and down with a frequency of 2.00 Hz. The maximum displacement of the end of the string is 20.00 cm. The first wave hits the lab wall 6.00 s after it was created. (a) What is the speed of the wave? (b) What is the period of the wave? (c) What is the wavelength of the wave?

Strategy

a. The speed of the wave can be derived by dividing the distance traveled by the time.
b. The period of the wave is the inverse of the frequency of the driving force.
c. The wavelength can be found from the speed and the period $v = \lambda / T$.

Solution

a. The first wave traveled 30.00 m in 6.00 s:
Equation:

$$v = \frac{30.00\,\text{m}}{6.00\,\text{s}} = 5.00\frac{\text{m}}{\text{s}}.$$

b. The period is equal to the inverse of the frequency:
Equation:

$$T = \frac{1}{f} = \frac{1}{2.00\,\text{s}^{-1}} = 0.50\,\text{s}.$$

c. The wavelength is equal to the velocity times the period:
Equation:

$$\lambda = vT = 5.00\frac{\text{m}}{\text{s}}(0.50\,\text{s}) = 2.50\,\text{m}.$$

Significance
The frequency of the wave produced by an oscillating driving force is equal to the frequency of the driving force.

Note:
Exercise:

Problem:

Check Your Understanding When a guitar string is plucked, the guitar string oscillates as a result of waves moving through the string. The vibrations of the string cause the air molecules to oscillate, forming sound waves. The frequency of the sound waves is equal to the frequency of the vibrating string. Is the wavelength of the sound wave always equal to the wavelength of the waves on the string?

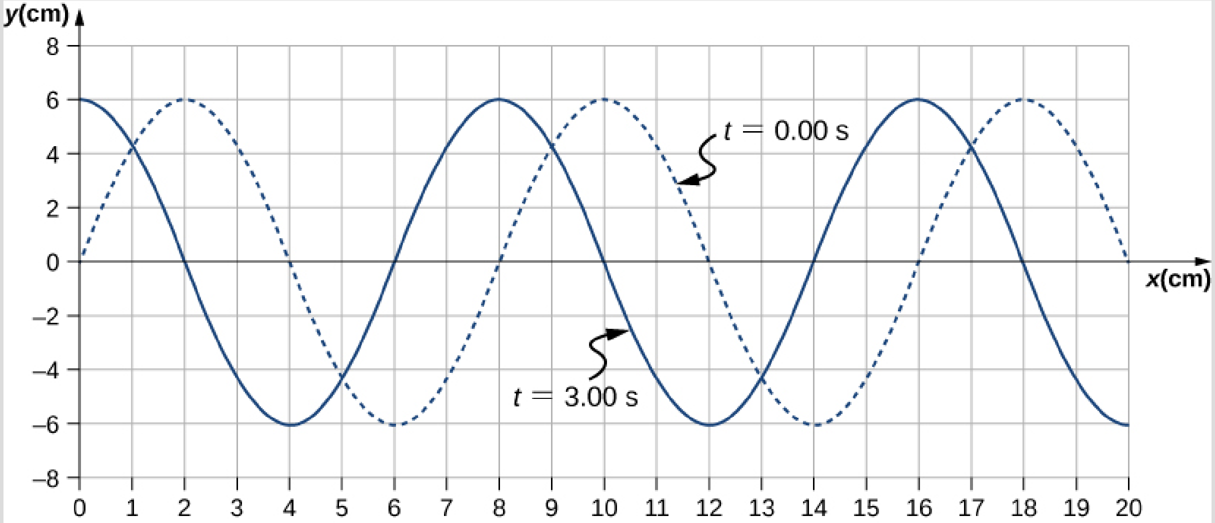
Solution:

The wavelength of the waves depends on the frequency and the velocity of the wave. The frequency of the sound wave is equal to the frequency of the wave on the string. The wavelengths of the sound waves and the waves on the string are equal only if the velocities of the waves are the same, which is not always the case. If the speed of the sound wave is different from the speed of the wave on the string, the wavelengths are different. This velocity of sound waves will be discussed in Sound.

Example:

Characteristics of a Wave

A transverse mechanical wave propagates in the positive *x*-direction through a spring (as shown in [link](a)) with a constant wave speed, and the medium oscillates between $+A$ and $-A$ around an equilibrium position. The graph in [link] shows the height of the spring (*y*) versus the position (*x*), where the *x*-axis points in the direction of propagation. The figure shows the height of the spring versus the *x*-position at $t = 0.00$ s as a dotted line and the wave at $t = 3.00$ s as a solid line. Assume the wave has not traveled more than 1 wavelength in this time. (a) Determine the wavelength and amplitude of the wave. (b) Find the propagation velocity of the wave. (c) Calculate the period and frequency of the wave.



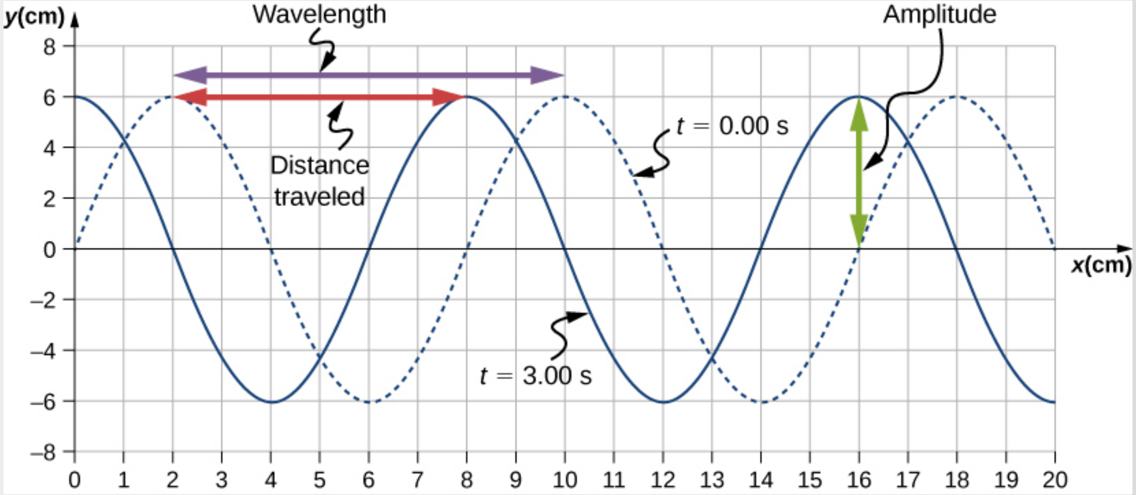
A transverse wave shown at two instants of time.

Strategy

a. The amplitude and wavelength can be determined from the graph.
b. Since the velocity is constant, the velocity of the wave can be found by dividing the distance traveled by the wave by the time it took the wave to travel the distance.
c. The period can be found from $v=\frac{\lambda}{T}$ and the frequency from $f=\frac{1}{T}$.

Solution

a. Read the wavelength from the graph, looking at the purple arrow in [link]. Read the amplitude by looking at the green arrow. The wavelength is $\lambda=8.00\text{ cm}$ and the amplitude is $A=6.00\text{ cm}$.



Characteristics of the wave marked on a graph of its displacement.

b. The distance the wave traveled from time $t=0.00\text{ s}$ to time $t=3.00\text{ s}$ can be seen in the graph. Consider the red arrow, which shows the distance the crest has moved in 3 s. The distance is $8.00\text{ cm}-2.00\text{ cm}=6.00\text{ cm}$. The velocity is
Equation:

$$v = \frac{\Delta x}{\Delta t} = \frac{8.00\ \text{cm} - 2.00\ \text{cm}}{3.00\ \text{s} - 0.00\ \text{s}} = 2.00\ \text{cm/s}.$$

c. The period is $T = \frac{\lambda}{v} = \frac{8.00\ \text{cm}}{2.00\ \text{cm/s}} = 4.00\ \text{s}$ and the frequency is $f = \frac{1}{T} = \frac{1}{4.00\ \text{s}} = 0.25\ \text{Hz}.$

Significance

Note that the wavelength can be found using any two successive identical points that repeat, having the same height and slope. You should choose two points that are most convenient. The displacement can also be found using any convenient point.

Note:

Exercise:

Problem:

Check Your Understanding The propagation velocity of a transverse or longitudinal mechanical wave may be constant as the wave disturbance moves through the medium. Consider a transverse mechanical wave: Is the velocity of the medium also constant?

Solution:

In a transverse wave, the wave may move at a constant propagation velocity through the medium, but the medium oscillates perpendicular to the motion of the wave. If the wave moves in the positive *x*-direction, the medium oscillates up and down in the *y*-direction. The velocity of the medium is therefore not constant, but the medium’s velocity and acceleration are similar to that of the simple harmonic motion of a mass on a spring.

Summary

- A wave is a disturbance that moves from the point of origin with a wave velocity v.
- A wave has a wavelength λ, which is the distance between adjacent identical parts of the wave. Wave velocity and wavelength are related to the wave's frequency and period by $v = \frac{\lambda}{T} = \lambda f$.
- Mechanical waves are disturbances that move through a medium and are governed by Newton's laws.
- Electromagnetic waves are disturbances in the electric and magnetic fields, and do not require a medium.
- Matter waves are a central part of quantum mechanics and are associated with protons, electrons, neutrons, and other fundamental particles found in nature.
- A transverse wave has a disturbance perpendicular to the wave's direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

Conceptual Questions

Exercise:

Problem:

Give one example of a transverse wave and one example of a longitudinal wave, being careful to note the relative directions of the disturbance and wave propagation in each.

Solution:

A wave on a guitar string is an example of a transverse wave. The disturbance of the string moves perpendicular to the propagation of the wave. The sound produced by the string is a longitudinal wave where the disturbance of the air moves parallel to the propagation of the wave.

Exercise:

Problem:

A sinusoidal transverse wave has a wavelength of 2.80 m. It takes 0.10 s for a portion of the string at a position *x* to move from a maximum position of $y = 0.03\,\text{m}$ to the equilibrium position $y = 0$. What are the period, frequency, and wave speed of the wave?

Exercise:

Problem:

What is the difference between propagation speed and the frequency of a mechanical wave? Does one or both affect wavelength? If so, how?

Solution:

Propagation speed is the speed of the wave propagating through the medium. If the wave speed is constant, the speed can be found by $v = \frac{\lambda}{T} = \lambda f$. The frequency is the number of wave that pass a point per unit time. The wavelength is directly proportional to the wave speed and inversely proportional to the frequency.

Exercise:

Problem:

Consider a stretched spring, such as a slinky. The stretched spring can support longitudinal waves and transverse waves. How can you produce transverse waves on the spring? How can you produce longitudinal waves on the spring?

Exercise:

Problem:

Consider a wave produced on a stretched spring by holding one end and shaking it up and down. Does the wavelength depend on the distance you move your hand up and down?

Solution:

No, the distance you move your hand up and down will determine the amplitude of the wave. The wavelength will depend on the frequency you move your hand up and down, and the speed of the wave through the spring.

Exercise:

Problem:

A sinusoidal, transverse wave is produced on a stretched spring, having a period *T*. Each section of the spring moves perpendicular to the direction of propagation of the wave, in simple harmonic motion with an amplitude *A*. Does each section oscillate with the same period as the wave or a different period? If the amplitude of the transverse wave were doubled but the period stays the same, would your answer be the same?

Exercise:

Problem:

An electromagnetic wave, such as light, does not require a medium. Can you think of an example that would support this claim?

Solution:

Light from the Sun and stars reach Earth through empty space where there is no medium present.

Problems

Exercise:

Problem:

Storms in the South Pacific can create waves that travel all the way to the California coast, 12,000 km away. How long does it take them to travel this distance if they travel at 15.0 m/s?

Exercise:

Problem:

Waves on a swimming pool propagate at 0.75 m/s. You splash the water at one end of the pool and observe the wave go to the opposite end, reflect, and return in 30.00 s. How far away is the other end of the pool?

Solution:

$$2d = vt \Rightarrow d = 11.25\,\text{m}$$

Exercise:

Problem:

Wind gusts create ripples on the ocean that have a wavelength of 5.00 cm and propagate at 2.00 m/s. What is their frequency?

Exercise:

Problem:

How many times a minute does a boat bob up and down on ocean waves that have a wavelength of 40.0 m and a propagation speed of 5.00 m/s?

Solution:

$$\begin{aligned} v &= f\lambda,\ \text{so that } f = 0.125\,\text{Hz, so that} \\ N &= 7.50\,\text{times} \end{aligned}$$

Exercise:

Problem:

Scouts at a camp shake the rope bridge they have just crossed and observe the wave crests to be 8.00 m apart. If they shake the bridge twice per second, what is the propagation speed of the waves?

Exercise:

Problem:

What is the wavelength of the waves you create in a swimming pool if you splash your hand at a rate of 2.00 Hz and the waves propagate at a wave speed of 0.800 m/s?

Solution:

$v = f\lambda \Rightarrow \lambda = 0.400\,\text{m}$

Exercise:

Problem:

What is the wavelength of an earthquake that shakes you with a frequency of 10.0 Hz and gets to another city 84.0 km away in 12.0 s?

Exercise:

Problem:

Radio waves transmitted through empty space at the speed of light $\left(v = c = 3.00\ \times\ 10^{8}\ \text{m/s}\right)$ by the *Voyager* spacecraft have a wavelength of 0.120 m. What is their frequency?

Solution:

$v = f\lambda \Rightarrow f = 2.50\ \times\ 10^{9}\ \text{Hz}$

Exercise:

Problem:

Your ear is capable of differentiating sounds that arrive at each ear just 0.34 ms apart, which is useful in determining where low frequency sound is originating from. (a) Suppose a low-frequency sound source is placed to the right of a person, whose ears are approximately 18 cm apart, and the speed of sound generated is 340 m/s. How long is the interval between when the sound arrives at the right ear and the sound arrives at the left ear? (b) Assume the same person was scuba diving and a low-frequency sound source was to the right of the scuba diver. How long is the interval between when the sound arrives at the right ear and the sound arrives at the left ear, if the speed of sound in water is 1500 m/s? (c) What is significant about the time interval of the two situations?

Exercise:

Problem:

(a) Seismographs measure the arrival times of earthquakes with a precision of 0.100 s. To get the distance to the epicenter of the quake, geologists compare the arrival times of S- and P-waves, which travel at different speeds. If S- and P-waves arrive at 4.00 and 7.20 km/s, respectively, in the region considered, how precisely can the distance to the source of the earthquake be determined? (b) Seismic waves from underground detonations of nuclear bombs can be used to locate the test site and detect violations of test bans. Discuss whether your answer to (a) implies a serious limit to such detection. (Note also that the uncertainty is greater if there is an uncertainty in the propagation speeds of the S- and P-waves.)

Solution:

a. The P-waves outrun the S-waves by a speed of $v = 3.20\,\text{km/s}$; therefore, $\Delta d = 0.320\,\text{km}$. b. Since the uncertainty in the distance is less than a kilometer, our answer to part (a) does not seem to limit the detection of nuclear bomb detonations. However, if the velocities are uncertain, then the uncertainty in the distance would increase and could then make it difficult to identify the source of the seismic waves.

Exercise:

Problem:

A Girl Scout is taking a 10.00-km hike to earn a merit badge. While on the hike, she sees a cliff some distance away. She wishes to estimate the time required to walk to the cliff. She knows that the speed of sound is approximately 343 meters per second. She yells and finds that the echo returns after approximately 2.00 seconds. If she can hike 1.00 km in 10 minutes, how long would it take her to reach the cliff?

Exercise:

Problem:

A quality assurance engineer at a frying pan company is asked to qualify a new line of nonstick-coated frying pans. The coating needs to be 1.00 mm thick. One method to test the thickness is for the engineer to pick a percentage of the pans manufactured, strip off the coating, and measure the thickness using a micrometer. This method is a destructive testing method. Instead, the engineer decides that every frying pan will be tested using a nondestructive method. An ultrasonic transducer is used that produces sound waves with a frequency of $f = 25\,\text{kHz}$. The sound waves are sent through the coating and are reflected by the interface between the coating and the metal pan, and the time is recorded. The wavelength of the ultrasonic waves in the coating is 0.076 m. What should be the time recorded if the coating is the correct thickness (1.00 mm)?

Solution:

$$\begin{aligned} v &= 1900\,\text{m/s} \\ \Delta t &= 1.05\,\mu\,\text{s} \end{aligned}$$

Glossary

longitudinal wave
wave in which the disturbance is parallel to the direction of propagation

mechanical wave
wave that is governed by Newton’s laws and requires a medium

transverse wave
wave in which the disturbance is perpendicular to the direction of propagation

wave
disturbance that moves from its source and carries energy

wave velocity
velocity at which the disturbance moves; also called the propagation velocity

wave speed
magnitude of the wave velocity

wavelength
distance between adjacent identical parts of a wave

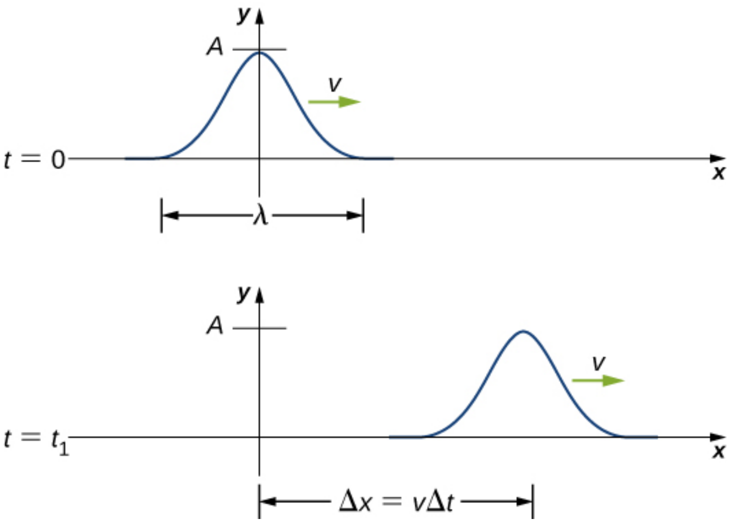
Mathematics of Waves
By the end of this section, you will be able to:

- Model a wave, moving with a constant wave velocity, with a mathematical expression
- Calculate the velocity and acceleration of the medium
- Show how the velocity of the medium differs from the wave velocity (propagation velocity)

In the previous section, we described periodic waves by their characteristics of wavelength, period, amplitude, and wave speed of the wave. Waves can also be described by the motion of the particles of the medium through which the waves move. The position of particles of the medium can be mathematically modeled as **wave functions**, which can be used to find the position, velocity, and acceleration of the particles of the medium of the wave at any time.

Pulses

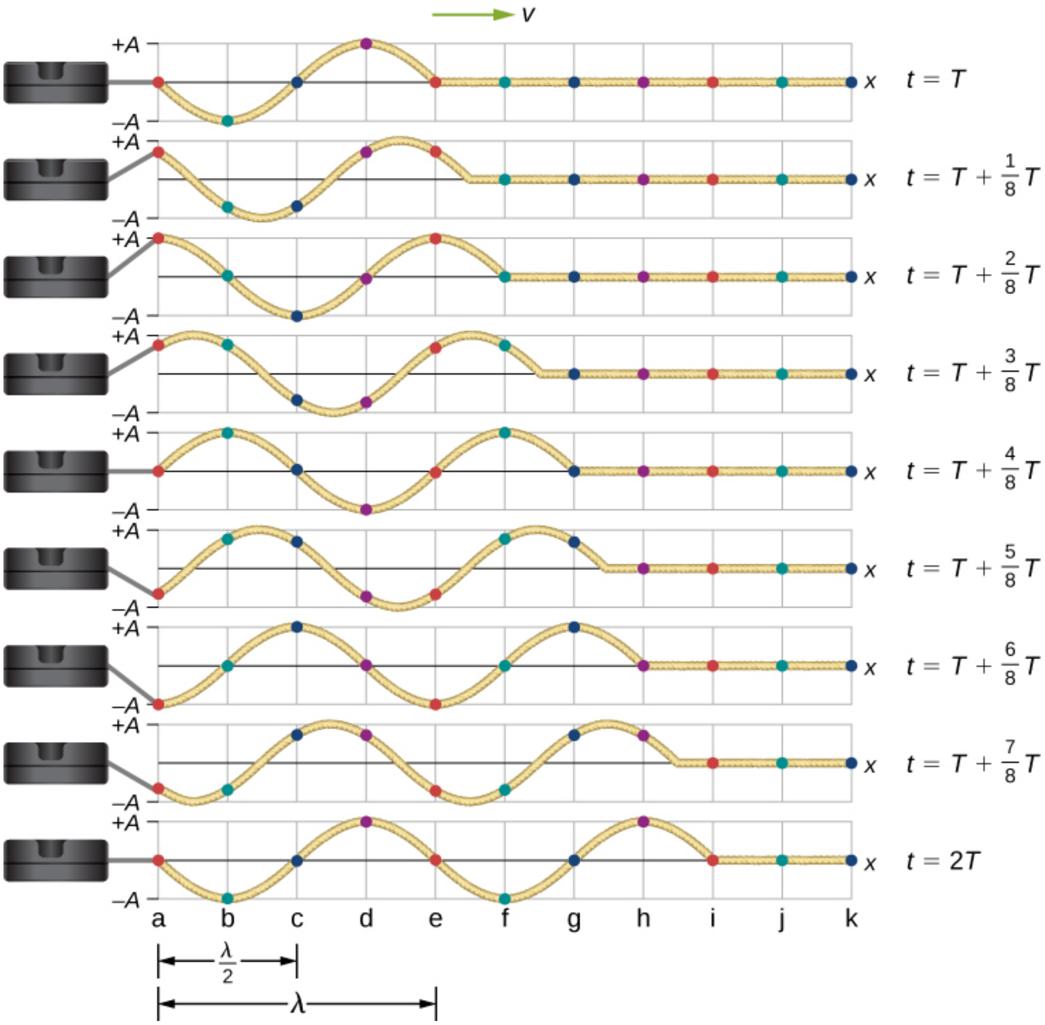
A **pulse** can be described as wave consisting of a single disturbance that moves through the medium with a constant amplitude. The pulse moves as a pattern that maintains its shape as it propagates with a constant wave speed. Because the wave speed is constant, the distance the pulse moves in a time Δt is equal to $\Delta x = v\Delta t$ ([link]).



The pulse at time $t = 0$ is centered on $x = 0$ with amplitude A. The pulse moves as a pattern with a constant shape, with a constant maximum value A. The velocity is constant and the pulse moves a distance $\Delta x = v\Delta t$ in a time Δt. The distance traveled is measured with any convenient point on the pulse. In this figure, the crest is used.

Modeling a One-Dimensional Sinusoidal Wave using a Wave Function

Consider a string kept at a constant tension F_T where one end is fixed and the free end is oscillated between $y = +A$ and $y = -A$ by a mechanical device at a constant frequency. [link] shows snapshots of the wave at an interval of an eighth of a period, beginning after one period $(t = T)$.



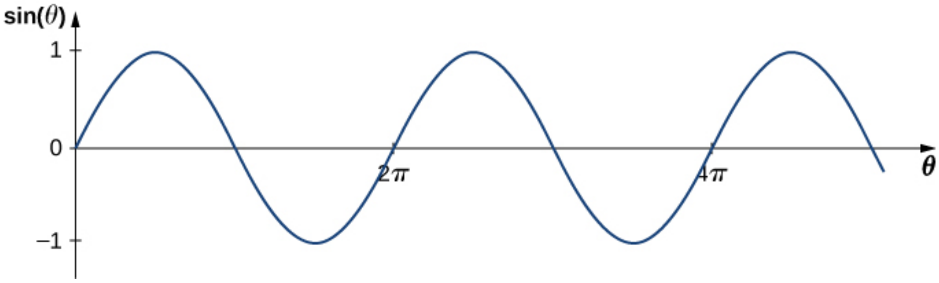
Snapshots of a transverse wave moving through a string under tension, beginning at time $t = T$ and taken at intervals of $\frac{1}{8}T$. Colored dots are used to highlight points on the string. Points that are a wavelength apart in the *x*-direction are highlighted with the same color dots.

Notice that each select point on the string (marked by colored dots) oscillates up and down in simple harmonic motion, between $y = +A$ and $y = -A$, with a period T. The wave on the string is sinusoidal and is translating in the positive *x*-direction as time progresses.

At this point, it is useful to recall from your study of algebra that if *f*(*x*) is some function, then $f\left(x - d\right)$ is the same function translated in the positive *x*-direction by a distance *d*. The function $f\left(x + d\right)$ is the same function translated in the negative *x*-direction by a distance *d*. We want to define a wave function that will give the *y*-position of each segment of the string for every position *x* along the string for every time *t*.

Looking at the first snapshot in [link], the *y*-position of the string between $x = 0$ and $x = \lambda$ can be modeled as a sine function. This wave propagates down the string one wavelength in one period, as seen in the last snapshot. The wave therefore moves with a constant wave speed of $v = \lambda / T$.

Recall that a sine function is a function of the angle θ, oscillating between $+1$ and -1, and repeating every 2π radians ([link]). However, the *y*-position of the medium, or the wave function, oscillates between $+A$ and $-A$, and repeats every wavelength λ.



A sine function oscillates between $+1$ and -1 every 2π radians.

To construct our model of the wave using a periodic function, consider the ratio of the angle and the position,

Equation:

$$\begin{aligned}\frac{\theta}{x} &= \frac{2\pi}{\lambda},\\ \theta &= \frac{2\pi}{\lambda}x.\end{aligned}$$

Using $\theta = \frac{2\pi}{\lambda}x$ and multiplying the sine function by the amplitude *A*, we can now model the *y*-position of the string as a function of the position *x*:

Equation:

$$y\left(x\right) = A\,\text{sin}\left(\frac{2\pi}{\lambda}x\right).$$

The wave on the string travels in the positive *x*-direction with a constant velocity *v*, and moves a distance *vt* in a time *t*. The wave function can now be defined by

Equation:

$$y\left(x,t\right) = A\,\text{sin}\left(\frac{2\pi}{\lambda}\left(x - vt\right)\right).$$

It is often convenient to rewrite this wave function in a more compact form. Multiplying through by the ratio $\frac{2\pi}{\lambda}$ leads to the equation

Equation:

$$y\left(x,t\right) = A\,\text{sin}\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{\lambda}vt\right).$$

The value $\frac{2\pi}{\lambda}$ is defined as the **wave number**. The symbol for the wave number is *k* and has units of inverse meters, m^{-1}:

Note:
Equation:

$$k \equiv \frac{2\pi}{\lambda}$$

Recall from Oscillations that the angular frequency is defined as $\omega \equiv \frac{2\pi}{T}$. The second term of the wave function becomes

Equation:

$$\frac{2\pi}{\lambda}vt = \frac{2\pi}{\lambda}\left(\frac{\lambda}{T}\right)t = \frac{2\pi}{T}t = \omega t.$$

The wave function for a simple harmonic wave on a string reduces to

Equation:

$$y\left(x, t\right) = A\,\text{sin}\left(kx \mp \omega t\right),$$

where A is the amplitude, $k = \frac{2\pi}{\lambda}$ is the wave number, $\omega = \frac{2\pi}{T}$ is the angular frequency, the minus sign is for waves moving in the positive *x*-direction, and the plus sign is for waves moving in the negative *x*-direction. The velocity of the wave is equal to

Note:
Equation:

$$v = \frac{\lambda}{T} = \frac{\lambda}{T}\left(\frac{2\pi}{2\pi}\right) = \frac{\omega}{k}.$$

Think back to our discussion of a mass on a spring, when the position of the mass was modeled as $x\left(t\right) = A\,\text{cos}\left(\omega t + \phi\right)$. The angle ϕ is a phase shift, added to allow for the fact that the mass may have initial conditions other than $x = +A$ and $v = 0$. For similar reasons, the initial phase is added to the wave function. The wave function modeling a sinusoidal wave, allowing for an initial phase shift ϕ, is

Note:
Equation:

$$y\left(x, t\right) = A\,\text{sin}\left(kx \mp \omega t + \phi\right)$$

The value

Note:
Equation:

$$\left(kx \mp \omega t + \phi\right)$$

is known as the phase of the wave, where ϕ is the initial phase of the wave function. Whether the temporal term ωt is negative or positive depends on the direction of the wave. First consider the minus sign for a wave with an initial phase equal to zero $(\phi = 0)$. The phase of the wave would be $(kx - \omega t)$. Consider following a point on a wave, such as a crest. A crest will occur when $\sin\,(kx - \omega t) = 1.00$, that is, when $kx - \omega t = n\pi + \frac{\pi}{2}$, for any integral value of *n*. For instance, one particular crest occurs at $kx - \omega t = \frac{\pi}{2}$. As the wave moves, time increases and *x* must also increase to keep the phase equal to $\frac{\pi}{2}$. Therefore, the minus sign is for a wave moving in the positive *x*-direction. Using the plus sign, $kx + \omega t = \frac{\pi}{2}$. As time increases, *x* must decrease to keep the phase equal to $\frac{\pi}{2}$. The plus sign is used for waves moving in the negative *x*-direction. In summary, $y\left(x, t\right) = A\,\sin\left(kx - \omega t + \phi\right)$ models a wave moving in the positive *x*-direction and $y\left(x, t\right) = A\,\sin\left(kx + \omega t + \phi\right)$ models a wave moving in the negative *x*-direction.

[link] is known as a simple harmonic wave function. A wave function is any function such that $f\left(x, t\right) = f\left(x - vt\right)$. Later in this chapter, we will see that it is a solution to the linear wave equation. Note that $y\left(x, t\right) = A\,\cos\left(kx + \omega t + \phi\prime\right)$ works equally well because it corresponds to a different phase shift $\phi\prime = \phi - \frac{\pi}{2}$.

Note:
Problem-Solving Strategy: Finding the Characteristics of a Sinusoidal Wave

1. To find the amplitude, wavelength, period, and frequency of a sinusoidal wave, write down the wave function in the form $y\left(x, t\right) = A\,\sin\left(kx - \omega t + \phi\right)$.
2. The amplitude can be read straight from the equation and is equal to *A*.
3. The period of the wave can be derived from the angular frequency $\left(T = \frac{2\pi}{\omega}\right)$.
4. The frequency can be found using $f = \frac{1}{T}$.
5. The wavelength can be found using the wave number $\left(\lambda = \frac{2\pi}{k}\right)$.

Example:
Characteristics of a Traveling Wave on a String
A transverse wave on a taut string is modeled with the wave function
Equation:

$$y\left(x, t\right) = A\,\sin\left(kx - wt\right) = 0.2\,\text{m}\,\sin\left(6.28\,\text{m}^{-1}x - 1.57\,\text{s}^{-1}t\right).$$

Find the amplitude, wavelength, period, and speed of the wave.
Strategy
All these characteristics of the wave can be found from the constants included in the equation or from simple combinations of these constants.
Solution

1. The amplitude, wave number, and angular frequency can be read directly from the wave equation:
   Equation:

$$y\left(x,t\right)=A\sin\left(kx-wt\right)=0.2\,\text{m}\,\sin\left(6.28\,\text{m}^{-1}x-1.57\,\text{s}^{-1}t\right).$$

Equation:

$$\left(A=0.2\,\text{m};\,k=6.28\,\text{m}^{-1};\,\omega=1.57\,\text{s}^{-1}\right)$$

2. The wave number can be used to find the wavelength:
   Equation:

$$\begin{array}{c}k=\frac{2\pi}{\lambda}.\\ \lambda=\frac{2\pi}{k}=\frac{2\pi}{6.28\,\text{m}^{-1}}=1.0\,\text{m}.\end{array}$$

3. The period of the wave can be found using the angular frequency:
   Equation:

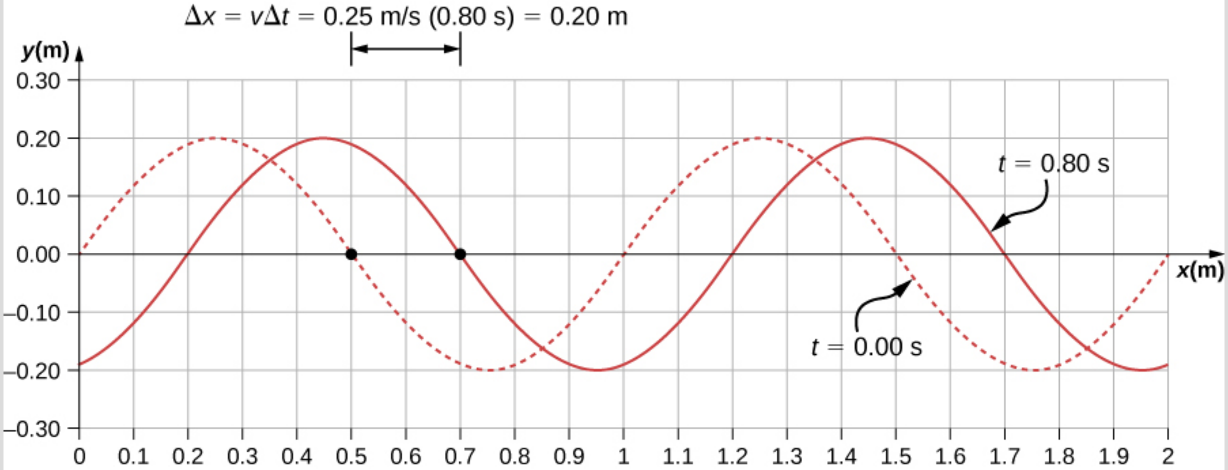
$$\begin{array}{c}\omega=\frac{2\pi}{T}.\\ T=\frac{2\pi}{\omega}=\frac{2\pi}{1.57\,\text{s}^{-1}}=4\,\text{s}.\end{array}$$

4. The speed of the wave can be found using the wave number and the angular frequency. The direction of the wave can be determined by considering the sign of $kx\mp\omega t$: A negative sign suggests that the wave is moving in the positive *x*-direction:
   Equation:

$$|v|=\frac{\omega}{k}=\frac{1.57\,\text{s}^{-1}}{6.28\,\text{m}^{-1}}=0.25\,\text{m/s}.$$

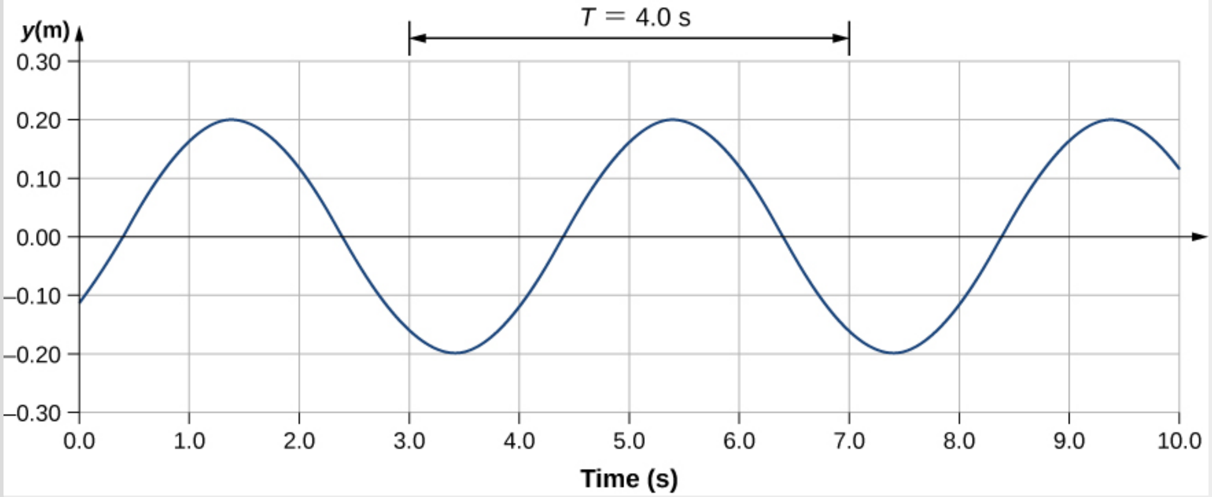
Significance

All of the characteristics of the wave are contained in the wave function. Note that the wave speed is the speed of the wave in the direction parallel to the motion of the wave. Plotting the height of the medium *y* versus the position *x* for two times $t=0.00\,\text{s}$ and $t=0.80\,\text{s}$ can provide a graphical visualization of the wave ([link]).



A graph of height of the wave *y* as a function of position *x* for snapshots of the wave at two times. The dotted line represents the wave at time $t=0.00\,\text{s}$ and the solid line represents the wave at $t=0.80\,\text{s}$. Since the wave velocity is constant, the distance the wave travels is the wave velocity times the time interval. The black dots indicate the points used to measure the displacement of the wave. The medium moves up and down, whereas the wave moves to the right.

There is a second velocity to the motion. In this example, the wave is transverse, moving horizontally as the medium oscillates up and down perpendicular to the direction of motion. The graph in [link] shows the motion of the medium at point $x = 0.60\,\text{m}$ as a function of time. Notice that the medium of the wave oscillates up and down between $y = +0.20\,\text{m}$ and $y = -0.20\,\text{m}$ every period of 4.0 seconds.



A graph of height of the wave *y* as a function of time *t* for the position $x = 0.6\,\text{m}$. The medium oscillates between $y = +0.20\,\text{m}$ and $y = -0.20\,\text{m}$ every period. The period represented picks two convenient points in the oscillations to measure the period. The period can be measured between any two adjacent points with the same amplitude and the same velocity, $(\partial y/\partial t)$. The velocity can be found by looking at the slope tangent to the point on a *y*-versus-*t* plot. Notice that at times $t = 3.00\ s$ and $t = 7.00s$, the heights and the velocities are the same and the period of the oscillation is 4.00 s.

Note:
Exercise:

Problem:

Check Your Understanding The wave function above is derived using a sine function. Can a cosine function be used instead?

Solution:

Yes, a cosine function is equal to a sine function with a phase shift, and either function can be used in a wave function. Which function is more convenient to use depends on the initial conditions. In [link], the wave has an initial height of $\text{y}(0.00, 0.00) = 0$ and then the wave height increases to the maximum height at the crest. If the initial height at the initial time was equal to the amplitude of the wave $\text{y}(0.00, 0.00) = +A$, then it might be more convenient to model the wave with a cosine function.

Velocity and Acceleration of the Medium

As seen in [link], the wave speed is constant and represents the speed of the wave as it propagates through the medium, not the speed of the particles that make up the medium. The particles of the medium oscillate around an equilibrium position as the wave propagates through the medium. In the case of the transverse wave propagating in the *x*-direction, the particles oscillate up and down in the *y*-direction, perpendicular to the motion of the wave. The velocity of the particles of the medium is not constant, which means there is an acceleration. The velocity of the medium, which is perpendicular to the wave velocity in a transverse wave, can be found by taking the partial derivative of the position equation with respect to time. The partial derivative is found by taking the derivative of the function, treating all variables as constants, except for the variable in question. In the case of the partial derivative with respect to time *t*, the position *x* is treated as a constant. Although this may sound strange if you haven't seen it before, the object of this exercise is to find the transverse velocity at a point, so in this sense, the *x*-position is not changing. We have

Equation:

$$\begin{aligned} y(x,t) &= A\sin(kx-\omega t+\phi) \\ v_y(x,t) &= \frac{\partial y(x,t)}{\partial t} = \frac{\partial}{\partial t}(A\sin(kx-\omega t+\phi)) \\ &= -A\omega\cos(kx-\omega t+\phi) \\ &= -v_{y\,\text{max}}\cos(kx-\omega t+\phi). \end{aligned}$$

The magnitude of the maximum velocity of the medium is $|v_{y_{\text{max}}}| = A\omega$. This may look familiar from the Oscillations and a mass on a spring.

We can find the acceleration of the medium by taking the partial derivative of the velocity equation with respect to time,

Equation:

$$\begin{aligned} a_y(x,t) &= \frac{\partial v_y}{\partial t} = \frac{\partial}{\partial t}(-A\omega\cos(kx-\omega t+\phi)) \\ &= -A\omega^2\sin(kx-\omega t+\phi) \\ &= -a_{y\,\text{max}}\sin(kx-\omega t+\phi). \end{aligned}$$

The magnitude of the maximum acceleration is $|a_{y_{\text{max}}}| = A\omega^2$. The particles of the medium, or the mass elements, oscillate in simple harmonic motion for a mechanical wave.

The Linear Wave Equation

We have just determined the velocity of the medium at a position *x* by taking the partial derivative, with respect to time, of the position *y*. For a transverse wave, this velocity is perpendicular to the direction of propagation of the wave. We found the acceleration by taking the partial derivative, with respect to time, of the velocity, which is the second time derivative of the position:

Equation:

$$a_y(x,t) = \frac{\partial^2 y(x,t)}{\partial t^2} = \frac{\partial^2}{\partial t^2}(A\sin(kx-\omega t+\phi)) = -A\omega^2\sin(kx-\omega t+\phi).$$

Now consider the partial derivatives with respect to the other variable, the position *x*, holding the time constant. The first derivative is the slope of the wave at a point *x* at a time *t*,

Equation:

$$\text{slope} = \frac{\partial y(x,t)}{\partial x} = \frac{\partial}{\partial x}(A\sin(kx-\omega t+\phi)) = Ak\cos(kx-\omega t+\phi).$$

The second partial derivative expresses how the slope of the wave changes with respect to position—in other words, the curvature of the wave, where

Equation:

$$\text{curvature} = \frac{\partial^2 y(x,t)}{\partial x^2} = \frac{\partial^2}{\partial^2 x}\left(A\sin(kx-\omega t+\phi)\right) = -Ak^2\sin(kx-\omega t+\phi).$$

The ratio of the acceleration and the curvature leads to a very important relationship in physics known as the **linear wave equation**. Taking the ratio and using the equation $v = \omega/k$ yields the linear wave equation (also known simply as the wave equation or the equation of a vibrating string),

Equation:

$$\begin{aligned}\frac{\frac{\partial^2 y(x,t)}{\partial t^2}}{\frac{\partial^2 y(x,t)}{\partial x^2}} &= \frac{-A\omega^2\sin(kx-\omega t+\phi)}{-Ak^2\sin(kx-\omega t+\phi)}\\ &= \frac{\omega^2}{k^2} = v^2,\end{aligned}$$

Note:

Equation:

$$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2}.$$

[link] is the linear wave equation, which is one of the most important equations in physics and engineering. We derived it here for a transverse wave, but it is equally important when investigating longitudinal waves. This relationship was also derived using a sinusoidal wave, but it successfully describes any wave or pulse that has the form $y(x,t) = f(x \mp vt)$. These waves result due to a linear restoring force of the medium—thus, the name linear wave equation. Any wave function that satisfies this equation is a linear wave function.

An interesting aspect of the linear wave equation is that if two wave functions are individually solutions to the linear wave equation, then the sum of the two linear wave functions is also a solution to the wave equation. Consider two transverse waves that propagate along the *x*-axis, occupying the same medium. Assume that the individual waves can be modeled with the wave functions $y_1(x,t) = f(x \mp vt)$ and $y_2(x,t) = g(x \mp vt)$, which are solutions to the linear wave equations and are therefore linear wave functions. The sum of the wave functions is the wave function

Equation:

$$y_1(x,t) + y_2(x,t) = f(x \mp vt) + g(x \mp vt).$$

Consider the linear wave equation:

Equation:

$$\begin{aligned}\frac{\partial^2 (f+g)}{\partial x^2} &= \frac{1}{v^2}\frac{\partial^2 (f+g)}{\partial t^2}\\ \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 g}{\partial x^2} &= \frac{1}{v^2}\left[\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 g}{\partial t^2}\right].\end{aligned}$$

This has shown that if two linear wave functions are added algebraically, the resulting wave function is also linear. This wave function models the displacement of the medium of the resulting wave at each position along the *x*-axis. If two linear waves occupy the same medium, they are said to interfere. If these waves can be modeled with a linear wave function, these wave functions add to form the wave equation of the wave resulting from the interference of the individual waves. The displacement of the medium at every point of the resulting wave is the algebraic sum of the displacements due to the individual waves.

Taking this analysis a step further, if wave functions $y_1(x,t) = f(x \mp vt)$ and $y_2(x,t) = g(x \mp vt)$ are solutions to the linear wave equation, then $Ay_1(x,t) + By_2(x,y)$, where *A* and *B* are constants, is also a solution to the linear wave equation. This property is known as the principle of superposition. Interference and superposition are covered in more detail in Interference of Waves.

Example:

Interference of Waves on a String

Consider a very long string held taut by two students, one on each end. Student A oscillates the end of the string producing a wave modeled with the wave function $y_1(x,t) = A\sin(kx - \omega t)$ and student B oscillates the string producing at twice the frequency, moving in the opposite direction. Both waves move at the same speed $v = \frac{\omega}{k}$. The two waves interfere to form a resulting wave whose wave function is $y_R(x,t) = y_1(x,t) + y_2(x,t)$. Find the velocity of the resulting wave using the linear wave equation $\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2}$.

Strategy

First, write the wave function for the wave created by the second student. Note that the angular frequency of the second wave is twice the frequency of the first wave (2ω), and since the velocity of the two waves are the same, the wave number of the second wave is twice that of the first wave $(2k)$. Next, write the wave equation for the resulting wave function, which is the sum of the two individual wave functions. Then find the second partial derivative with respect to position and the second partial derivative with respect to time. Use the linear wave equation to find the velocity of the resulting wave.

Solution

1. Write the wave function of the second wave: $y_2(x,t) = A\sin(2kx + 2\omega t)$.
2. Write the resulting wave function:
   Equation:

$$y_R(x,t) = y_1(x,t) + y(x,t) = A\sin(kx - \omega t) + A\sin(2kx + 2\omega t).$$

3. Find the partial derivatives:
   Equation:

$$\begin{aligned}
\frac{\partial y_R(x,t)}{\partial x} &= -Ak\cos(kx - \omega t) + 2Ak\cos(2kx + 2\omega t), \\
\frac{\partial^2 y_R(x,t)}{\partial^2 x} &= -Ak^2\sin(kx - \omega t) - 4Ak^2\sin(2kx + 2\omega t), \\
\frac{\partial y_R(x,t)}{\partial t} &= -A\omega\cos(kx - \omega t) + 2A\omega\cos(2kx + 2\omega t), \\
\frac{\partial^2 y_R(x,t)}{\partial^2 t} &= -A\omega^2\sin(kx - \omega t) - 4A\omega^2\sin(2kx + 2\omega t).
\end{aligned}$$

4. Use the wave equation to find the velocity of the resulting wave:

$$\begin{aligned}\frac{\partial^2 y(x,t)}{\partial x^2} &= \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2},\\ -Ak^2\sin(kx-\omega t) - 4Ak^2\sin(2kx+2\omega t) &= \frac{1}{v^2}\left(-A\omega^2\sin(kx-\omega t) - 4A\omega^2\sin(2kx+2\omega t)\right),\\ k^2\left(-A\sin(kx-\omega t) - 4A\sin(2kx+2\omega t)\right) &= \frac{\omega^2}{v^2}\left(-A\sin(kx-\omega t) - 4A\sin(2kx+2\omega t)\right),\\ k^2 &= \frac{\omega^2}{v^2},\ |v| = \frac{\omega}{k}.\end{aligned}$$

Significance
The speed of the resulting wave is equal to the speed of the original waves $\left(v = \frac{\omega}{k}\right)$. We will show in the next section that the speed of a simple harmonic wave on a string depends on the tension in the string and the mass per length of the string. For this reason, it is not surprising that the component waves as well as the resultant wave all travel at the same speed.

Note:
Exercise:

Problem:

Check Your Understanding The wave equation $\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2}$ works for any wave of the form $y(x,t) = f(x \mp vt)$. In the previous section, we stated that a cosine function could also be used to model a simple harmonic mechanical wave. Check if the wave
Equation:

$$y(x,t) = 0.50\,\text{m}\cos\left(0.20\pi\,\text{m}^{-1}x - 4.00\pi\,\text{s}^{-1}t + \frac{\pi}{10}\right)$$

is a solution to the wave equation.

Solution:

This wave, with amplitude $A = 0.5\,\text{m}$, wavelength $\lambda = 10.00\,\text{m}$, period $T = 0.50\,\text{s}$, is a solution to the wave equation with a wave velocity $v = 20.00\,\text{m/s}$.

Any disturbance that complies with the wave equation can propagate as a wave moving along the *x*-axis with a wave speed *v*. It works equally well for waves on a string, sound waves, and electromagnetic waves. This equation is extremely useful. For example, it can be used to show that electromagnetic waves move at the speed of light.

Summary

- A wave is an oscillation (of a physical quantity) that travels through a medium, accompanied by a transfer of energy. Energy transfers from one point to another in the direction of the wave motion. The particles of the medium oscillate up and down, back and forth, or both up and down and back and forth, around an equilibrium position.
- A snapshot of a sinusoidal wave at time $t = 0.00\,\text{s}$ can be modeled as a function of position. Two examples of such functions are $y(x) = A\sin(kx+\phi)$ and $y(x) = A\cos(kx+\phi)$.
- Given a function of a wave that is a snapshot of the wave, and is only a function of the position *x*, the motion of the pulse or wave moving at a constant velocity can be modeled with the function, replacing *x* with $x \mp vt$. The minus sign is for motion in the positive direction and the plus sign for the negative direction.

- The wave function is given by $y\left(x,t\right)=A\sin\left(kx-\omega t+\phi\right)$ where $k=2\pi/\lambda$ is defined as the wave number, $\omega=2\pi/T$ is the angular frequency, and ϕ is the phase shift.
- The wave moves with a constant velocity v_w, where the particles of the medium oscillate about an equilibrium position. The constant velocity of a wave can be found by $v=\frac{\lambda}{T}=\frac{\omega}{k}$.

Conceptual Questions

Exercise:

Problem:

If you were to shake the end of a taut spring up and down 10 times a second, what would be the frequency and the period of the sinusoidal wave produced on the spring?

Exercise:

Problem:

If you shake the end of a stretched spring up and down with a frequency *f*, you can produce a sinusoidal, transverse wave propagating down the spring. Does the wave number depend on the frequency you are shaking the spring?

Solution:

The wavelength is equal to the velocity of the wave times the frequency and the wave number is equal to $k=\frac{2\pi}{\lambda}$, so yes, the wave number will depend on the frequency and also depend on the velocity of the wave propagating through the spring.

Exercise:

Problem:

Does the vertical speed of a segment of a horizontal taut string through which a sinusoidal, transverse wave is propagating depend on the wave speed of the transverse wave?

Exercise:

Problem:

In this section, we have considered waves that move at a constant wave speed. Does the medium accelerate?

Solution:

The medium moves in simple harmonic motion as the wave propagates through the medium, continuously changing speed, therefore it accelerates. The acceleration of the medium is due to the restoring force of the medium, which acts in the opposite direction of the displacement.

Exercise:

Problem:

If you drop a pebble in a pond you may notice that several concentric ripples are produced, not just a single ripple. Why do you think that is?

Problems

Exercise:

Problem:

A pulse can be described as a single wave disturbance that moves through a medium. Consider a pulse that is defined at time $t = 0.00\,\text{s}$ by the equation $y\left(x\right) = \frac{6.00\,\text{m}^3}{x^2 + 2.00\,\text{m}^2}$ centered around $x = 0.00\,\text{m}$. The pulse moves with a velocity of $v = 3.00\,\text{m/s}$ in the positive *x*-direction. (a) What is the amplitude of the pulse? (b) What is the equation of the pulse as a function of position and time? (c) Where is the pulse centered at time $t = 5.00\,\text{s}$?

Exercise:

Problem:

A transverse wave on a string is modeled with the wave function $y\left(x,t\right) = \left(0.20\,\text{cm}\right)\text{sin}\left(2.00\,\text{m}^{-1}x - 3.00\,\text{s}^{-1}t + \frac{\pi}{16}\right)$. What is the height of the string with respect to the equilibrium position at a position $x = 4.00\,\text{m}$ and a time $t = 10.00\,\text{s}$?

Solution:

$y\left(x,t\right) = -0.037\,\text{cm}$

Exercise:

Problem:

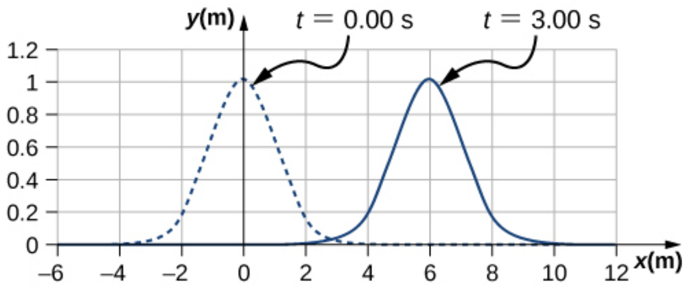
Consider the wave function $y\left(x,t\right) = \left(3.00\,\text{cm}\right)\text{sin}\left(0.4\,\text{m}^{-1}x + 2.00\,\text{s}^{-1}t + \frac{\pi}{10}\right)$. What are the period, wavelength, speed, and initial phase shift of the wave modeled by the wave function?

Exercise:

Problem:

A pulse is defined as $y\left(x,t\right) = e^{-2.77\left(\frac{2.00\left(x - 2.00\,\text{m/s}\left(t\right)\right)}{5.00\,\text{m}}\right)^2}$. Use a spreadsheet, or other computer program, to plot the pulse as the height of medium *y* as a function of position *x*. Plot the pulse at times $t = 0.00\,\text{s}$ and $t = 3.00\,\text{s}$ on the same graph. Where is the pulse centered at time $t = 3.00\,\text{s}$? Use your spreadsheet to check your answer.

Solution:



The pulse will move $\Delta x = 6.00\,\text{m}$.

Exercise:

Problem:

A wave is modeled at time $t = 0.00\,\text{s}$ with a wave function that depends on position. The equation is $y\left(x\right) = \left(0.30\,\text{m}\right)\text{sin}\left(6.28\,\text{m}^{-1}x\right)$. The wave travels a distance of 4.00 meters in 0.50 s in the positive *x*-direction. Write an equation for the wave as a function of position and time.

Exercise:

Problem:

A wave is modeled with the function $y\left(x,t\right)=\left(0.25\ \text{m}\right)\text{cos}\left(0.30\ \text{m}^{-1}x-0.90\ \text{s}^{-1}t+\frac{\pi}{3}\right)$. Find the (a) amplitude, (b) wave number, (c) angular frequency, (d) wave speed, (e) initial phase shift, (f) wavelength, and (g) period of the wave.

Solution:

a. $A=0.25\ \text{m}$; b. $k=0.30\ \text{m}^{-1}$; c. $\omega=0.90\ \text{s}^{-1}$; d. $v=3.0\ \text{m/s}$; e. $\phi=\pi/3\ \text{rad}$; f. $\lambda=20.93\ \text{m}$; g. $T=6.98\ \text{s}$

Exercise:

Problem:

A surface ocean wave has an amplitude of 0.60 m and the distance from trough to trough is 8.00 m. It moves at a constant wave speed of 1.50 m/s propagating in the positive *x*-direction. At $t=0$, the water displacement at $x=0$ is zero, and v_y is positive. (a) Assuming the wave can be modeled as a sine wave, write a wave function to model the wave. (b) Use a spreadsheet to plot the wave function at times $t=0.00\ \text{s}$ and $t=2.00\ \text{s}$ on the same graph. Verify that the wave moves 3.00 m in those 2.00 s.

Exercise:

Problem:

A wave is modeled by the wave function $y\left(x,t\right)=\left(0.30\ \text{m}\right)\text{sin}\left[\frac{2\pi}{4.50\ \text{m}}\left(x-18.00\frac{\text{m}}{\text{s}}t\right)\right]$. What are the amplitude, wavelength, wave speed, period, and frequency of the wave?

Solution:

$A=0.30\ \text{m},\ \lambda=4.50\ \text{m},\ v=18.00\ \text{m/s},\ f=4.00\ \text{Hz},\ T=0.25\ \text{s}$

Exercise:

Problem:

A transverse wave on a string is described with the wave function $y\left(x,t\right)=\left(0.50\ \text{cm}\right)\text{sin}\left(1.57\ \text{m}^{-1}x-6.28\ \text{s}^{-1}t\right)$. (a) What is the wave velocity of the wave? (b) What is the magnitude of the maximum velocity of the string perpendicular to the direction of the motion?

Exercise:

Problem:

A swimmer in the ocean observes one day that the ocean surface waves are periodic and resemble a sine wave. The swimmer estimates that the vertical distance between the crest and the trough of each wave is approximately 0.45 m, and the distance between each crest is approximately 1.8 m. The swimmer counts that 12 waves pass every two minutes. Determine the simple harmonic wave function that would describes these waves.

Solution:

$y\left(x,t\right)=0.23\ \text{m}\sin\left(3.49\ \text{m}^{-1}x-0.63\ \text{s}^{-1}t\right)$

Exercise:

Problem:

Consider a wave described by the wave function $y\left(x,t\right)=0.3\,\text{m}\,\text{sin}\left(2.00\,{\text{m}}^{-1}x-628.00\,{\text{s}}^{-1}t\right).$ (a) How many crests pass by an observer at a fixed location in 2.00 minutes? (b) How far has the wave traveled in that time?

Exercise:

Problem:

Consider two waves defined by the wave functions ${y}_{1}\left(x,t\right)=0.50\,\text{m}\,\text{sin}\left(\frac{2\pi }{3.00\,\text{m}}x+\frac{2\pi }{4.00\,\text{s}}t\right)$ and ${y}_{2}\left(x,t\right)=0.50\,\text{m}\,\text{sin}\left(\frac{2\pi }{6.00\,\text{m}}x-\frac{2\pi }{4.00\,\text{s}}t\right).$ What are the similarities and differences between the two waves?

Solution:

They have the same angular frequency, frequency, and period. They are traveling in opposite directions and ${y}_{2}\left(x,t\right)$ has twice the wavelength as ${y}_{1}\left(x,t\right)$ and is moving at half the wave speed.

Exercise:

Problem:

Consider two waves defined by the wave functions ${y}_{1}\left(x,t\right)=0.20\,\text{m}\,\text{sin}\left(\frac{2\pi }{6.00\,\text{m}}x-\frac{2\pi }{4.00\,\text{s}}t\right)$ and ${y}_{2}\left(x,t\right)=0.20\,\text{m}\,\text{cos}\left(\frac{2\pi }{6.00\,\text{m}}x-\frac{2\pi }{4.00\,\text{s}}t\right).$ What are the similarities and differences between the two waves?

Exercise:

Problem:

The speed of a transverse wave on a string is 300.00 m/s, its wavelength is 0.50 m, and the amplitude is 20.00 cm. How much time is required for a particle on the string to move through a distance of 5.00 km?

Solution:

Each particle of the medium moves a distance of 4*A* each period. The period can be found by dividing the velocity by the wavelength: $t=10.42\,\text{s}$

Glossary

linear wave equation
equation describing waves that result from a linear restoring force of the medium; any function that is a solution to the wave equation describes a wave moving in the positive *x*-direction or the negative *x*-direction with a constant wave speed *v*

pulse
single disturbance that moves through a medium, transferring energy but not mass

wave function
mathematical model of the position of particles of the medium

wave number
$\frac{2\pi }{\lambda }$

Energy and Power of a Wave

By the end of this section, you will be able to:

- Explain how energy travels with a pulse or wave
- Describe, using a mathematical expression, how the energy in a wave depends on the amplitude of the wave

All waves carry energy, and sometimes this can be directly observed. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls ([link]). Loud sounds can pulverize nerve cells in the inner ear, causing permanent hearing loss. Ultrasound is used for deep-heat treatment of muscle strains. A laser beam can burn away a malignancy. Water waves chew up beaches.

The destructive effect of an earthquake is observable evidence of the energy carried in these waves. The Richter scale rating of earthquakes is a logarithmic scale related to both their amplitude and the energy they carry.

In this section, we examine the quantitative expression of energy in waves. This will be of fundamental importance in later discussions of waves, from sound to light to quantum mechanics.

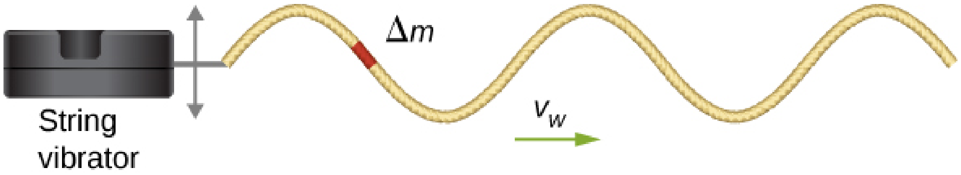
Energy in Waves

The amount of energy in a wave is related to its amplitude and its frequency. Large-amplitude earthquakes produce large ground displacements. Loud sounds have high-pressure amplitudes and come from larger-amplitude source vibrations than soft sounds. Large ocean breakers churn up the shore more than small ones. Consider the example of the seagull and the water wave earlier in the chapter ([link]). Work is done on the seagull by the wave as the seagull is moved up, changing its potential energy. The larger the amplitude, the higher the seagull is lifted by the wave and the larger the change in potential energy.

The energy of the wave depends on both the amplitude and the frequency. If the energy of each wavelength is considered to be a discrete packet of energy, a high-frequency wave will deliver more of these packets per unit time than a low-frequency wave. We will see that the average rate of energy transfer in mechanical waves is proportional to both the square of the amplitude and the square of the frequency. If two mechanical waves have equal amplitudes, but one wave has a frequency equal to twice the frequency of the other, the higher-frequency wave will have a rate of energy transfer a factor of four times as great as the rate of energy transfer of the lower-frequency wave. It should be noted that although the rate of energy transport is proportional to both the square of the amplitude and square of the frequency in mechanical waves, the rate of energy transfer in electromagnetic waves is proportional to the square of the amplitude, but independent of the frequency.

Power in Waves

Consider a sinusoidal wave on a string that is produced by a string vibrator, as shown in [link]. The string vibrator is a device that vibrates a rod up and down. A string of uniform linear mass density is attached to the rod, and the rod oscillates the string, producing a sinusoidal wave. The rod does work on the string, producing energy that propagates along the string. Consider a mass element of the string with a mass Δm, as seen in [link]. As the energy propagates along the string, each mass element of the string is driven up and down at the same frequency as the wave. Each mass element of the string can be modeled as a simple harmonic oscillator. Since the string has a constant linear density $\mu = \frac{\Delta m}{\Delta x}$, each mass element of the string has the mass $\Delta m = \mu \Delta x$.



A string vibrator is a device that vibrates a rod. A string is attached to the rod, and the rod does work on the string, driving the string up and down. This produces a sinusoidal wave in the string, which moves with a wave velocity *v*. The wave speed depends on the tension in the string and the linear mass density of the string. A section of the string with mass Δm oscillates at the same frequency as the wave.

The total mechanical energy of the wave is the sum of its kinetic energy and potential energy. The kinetic energy $K = \frac{1}{2}mv^2$ of each mass element of the string of length Δx is $\Delta K = \frac{1}{2}(\Delta m)v_y^2$, as the mass element oscillates perpendicular to the direction of the motion of the wave. Using the constant linear mass density, the kinetic energy of each mass element of the string with length Δx is

Equation:

$$\Delta K = \frac{1}{2}(\mu \Delta x)v_y^2.$$

A differential equation can be formed by letting the length of the mass element of the string approach zero,

Equation:

$$dK = \lim_{\Delta x \to 0} \frac{1}{2}(\mu \Delta x)v_y^2 = \frac{1}{2}(\mu dx)v_y^2.$$

Since the wave is a sinusoidal wave with an angular frequency ω, the position of each mass element may be modeled as $y\left(x, t\right) = A \sin\left(kx - \omega t\right)$. Each mass element of the string oscillates with a velocity $v_y = \frac{\partial y(x,t)}{\partial t} = -A\omega \cos\left(kx - \omega t\right)$. The kinetic energy of each mass element of the string becomes

Equation:

$$\begin{aligned} dK &= \tfrac{1}{2}(\mu dx)(-A\omega \cos\left(kx - \omega t\right))^2, \\ &= \tfrac{1}{2}(\mu dx)A^2\omega^2 \cos^2\left(kx - \omega t\right). \end{aligned}$$

The wave can be very long, consisting of many wavelengths. To standardize the energy, consider the kinetic energy associated with a wavelength of the wave. This kinetic energy can be integrated over the wavelength to find the energy associated with each wavelength of the wave:

Equation:

$$\begin{aligned} dK &= \tfrac{1}{2}(\mu dx)A^2\omega^2\cos^2(kx), \\ \int_0^{K_\lambda} dK &= \int_0^{\lambda} \frac{1}{2}\mu A^2\omega^2\cos^2(kx)dx = \tfrac{1}{2}\mu A^2\omega^2 \int_0^{\lambda}\cos^2(kx)dx, \\ K_\lambda &= \tfrac{1}{2}\mu A^2\omega^2\big[\tfrac{1}{2}x + \tfrac{1}{4k}\sin(2kx)\big]_0^{\lambda} = \tfrac{1}{2}\mu A^2\omega^2\left[\tfrac{1}{2}\lambda + \tfrac{1}{4k}\sin(2k\lambda) - \tfrac{1}{4k}\sin(0)\right], \\ K_\lambda &= \tfrac{1}{4}\mu A^2\omega^2\lambda. \end{aligned}$$

There is also potential energy associated with the wave. Much like the mass oscillating on a spring, there is a conservative restoring force that, when the mass element is displaced from the equilibrium position, drives the mass element back to the equilibrium position. The potential energy of the mass element can be found by considering the linear restoring force of the string, In Oscillations, we saw that the potential energy stored in a spring with a linear restoring force is equal to $U = \frac{1}{2}k_s x^2$, where the equilibrium position is defined as $x = 0.00\,\text{m}$. When a mass attached to the spring oscillates in simple harmonic motion, the angular frequency is equal to $\omega = \sqrt{\frac{k_s}{m}}$. As each mass element oscillates in simple harmonic motion, the spring constant is equal to $k_s = \Delta m\omega^2$. The potential energy of the mass element is equal to

Equation:

$$\Delta U = \frac{1}{2}k_s x^2 = \frac{1}{2}\Delta m\omega^2 x^2.$$

Note that k_s is the spring constant and not the wave number $k = \frac{2\pi}{\lambda}$. This equation can be used to find the energy over a wavelength. Integrating over the wavelength, we can compute the potential energy over a wavelength:

Equation:

$$\begin{aligned} dU &= \tfrac{1}{2}k_s x^2 = \tfrac{1}{2}\mu\omega^2 x^2 dx, \\ U_\lambda &= \tfrac{1}{2}\mu\omega^2 A^2 \int_0^{\lambda}\sin^2(kx)dx = \tfrac{1}{4}\mu A^2\omega^2\lambda. \end{aligned}$$

The potential energy associated with a wavelength of the wave is equal to the kinetic energy associated with a wavelength.

The total energy associated with a wavelength is the sum of the potential energy and the kinetic energy:

Equation:

$$\begin{aligned}E_\lambda &= U_\lambda + K_\lambda,\\ E_\lambda &= \tfrac{1}{4}\mu A^2\omega^2\lambda + \tfrac{1}{4}\mu A^2\omega^2\lambda = \tfrac{1}{2}\mu A^2\omega^2\lambda.\end{aligned}$$

The time-averaged power of a sinusoidal mechanical wave, which is the average rate of energy transfer associated with a wave as it passes a point, can be found by taking the total energy associated with the wave divided by the time it takes to transfer the energy. If the velocity of the sinusoidal wave is constant, the time for one wavelength to pass by a point is equal to the period of the wave, which is also constant. For a sinusoidal mechanical wave, the time-averaged power is therefore the energy associated with a wavelength divided by the period of the wave. The wavelength of the wave divided by the period is equal to the velocity of the wave,

Note:
Equation:

$$P_{\text{ave}} = \frac{E_\lambda}{T} = \frac{1}{2}\mu A^2\omega^2\frac{\lambda}{T} = \frac{1}{2}\mu A^2\omega^2 v.$$

Note that this equation for the time-averaged power of a sinusoidal mechanical wave shows that the power is proportional to the square of the amplitude of the wave and to the square of the angular frequency of the wave. Recall that the angular frequency is equal to $\omega = 2\pi f$, so the power of a mechanical wave is equal to the square of the amplitude and the square of the frequency of the wave.

Example:
Power Supplied by a String Vibrator
Consider a two-meter-long string with a mass of 70.00 g attached to a string vibrator as illustrated in [link]. The tension in the string is 90.0 N. When the string vibrator is turned on, it oscillates with a frequency of 60 Hz and produces a sinusoidal wave on the string with an amplitude of 4.00 cm and a constant wave speed. What is the time-averaged power supplied to the wave by the string vibrator?

Strategy
The power supplied to the wave should equal the time-averaged power of the wave on the string. We know the mass of the string (m_s), the length of the string (L_s), and the tension (F_T) in the string. The speed of the wave on the string can be derived from the linear mass density and the tension. The string oscillates with the same frequency as the string vibrator, from which we can find the angular frequency.

Solution

1. Begin with the equation of the time-averaged power of a sinusoidal wave on a string:
   Equation:

   $$P=\frac{1}{2}\mu A^2\omega^2 v.$$

   The amplitude is given, so we need to calculate the linear mass density of the string, the angular frequency of the wave on the string, and the speed of the wave on the string.
2. We need to calculate the linear density to find the wave speed:
   Equation:

   $$\mu=\frac{m_s}{L_s}=\frac{0.070\,\text{kg}}{2.00\,\text{m}}=0.035\,\text{kg/m}.$$

3. The wave speed can be found using the linear mass density and the tension of the string:
   Equation:

   $$v=\sqrt{\frac{F_T}{\mu}}=\sqrt{\frac{90.00\,\text{N}}{0.035\,\text{kg/m}}}=50.71\,\text{m/s}.$$

4. The angular frequency can be found from the frequency:
   Equation:

   $$\omega=2\pi f=2\pi\left(60\,\text{s}^{-1}\right)=376.80\,\text{s}^{-1}.$$

5. Calculate the time-averaged power:
   Equation:

   $$P=\frac{1}{2}\mu A^2\omega^2 v=\frac{1}{2}\left(0.035\frac{\text{kg}}{\text{m}}\right)(0.040\,\text{m})^2\left(376.80\,\text{s}^{-1}\right)^2\left(50.71\frac{\text{m}}{\text{s}}\right)=201.59\,\text{W}.$$

Significance
The time-averaged power of a sinusoidal wave is proportional to the square of the amplitude of the wave and the square of the angular frequency of the wave. This is true for most mechanical waves. If either the angular frequency or the amplitude of the wave were doubled, the power would increase by a factor of four. The time-averaged power of the wave on a string is also proportional to the speed of the sinusoidal wave on the string. If the speed were doubled, by increasing the tension by a factor of four, the power would also be doubled.

Note:
Exercise:

Problem:

Check Your Understanding Is the time-averaged power of a sinusoidal wave on a string proportional to the linear density of the string?

Solution:

At first glance, the time-averaged power of a sinusoidal wave on a string may look proportional to the linear density of the string because $P = \frac{1}{2}\mu A^2\omega^2 v$; however, the speed of the wave depends on the linear density. Replacing the wave speed with $\sqrt{\frac{F_T}{\mu}}$ shows that the power is proportional to the square root of tension and proportional to the square root of the linear mass density:
$P = \frac{1}{2}\mu A^2\omega^2 v = \frac{1}{2}\mu A^2\omega^2\sqrt{\frac{F_T}{\mu}} = \frac{1}{2}A^2\omega^2\sqrt{\mu F_T}.$

The equations for the energy of the wave and the time-averaged power were derived for a sinusoidal wave on a string. In general, the energy of a mechanical wave and the power are proportional to the amplitude squared and to the angular frequency squared (and therefore the frequency squared).

Another important characteristic of waves is the intensity of the waves. Waves can also be concentrated or spread out. Waves from an earthquake, for example, spread out over a larger area as they move away from a source, so they do less damage the farther they get from the source. Changing the area the waves cover has important effects. All these pertinent factors are included in the definition of **intensity (I)** as power per unit area:

Note:
Equation:

$$I = \frac{P}{A},$$

where P is the power carried by the wave through area A. The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter (W/m^2). Many waves are spherical waves that move out from a source as a sphere. For example, a sound speaker mounted on a post above the ground may produce sound waves that move away from the source as a spherical wave. Sound waves are discussed in more detail in the next chapter, but in general, the farther you are from the speaker, the less intense the sound you hear. As a spherical wave moves out from a source, the surface area of the wave increases as the radius increases $\left(A = 4\pi r^2\right)$. The intensity for a spherical wave is therefore

Note:
Equation:

$$I = \frac{P}{4\pi r^2}.$$

If there are no dissipative forces, the energy will remain constant as the spherical wave moves away from the source, but the intensity will decrease as the surface area increases.

In the case of the two-dimensional circular wave, the wave moves out, increasing the circumference of the wave as the radius of the circle increases. If you toss a pebble in a pond, the surface ripple moves out as a circular wave. As the ripple moves away from the source, the amplitude decreases. The energy of the wave spreads around a larger circumference and the amplitude decreases proportional to $\frac{1}{r}$, which is also the same in the case of a spherical wave, since intensity is proportional to the amplitude squared.

Summary

- The energy and power of a wave are proportional to the square of the amplitude of the wave and the square of the angular frequency of the wave.
- The time-averaged power of a sinusoidal wave on a string is found by $P_{\text{ave}} = \frac{1}{2}\mu A^2\omega^2 v$, where μ is the linear mass density of the string, A is the amplitude of the wave, ω is the angular frequency of the wave, and v is the speed of the wave.
- Intensity is defined as the power divided by the area. In a spherical wave, the area is $A = 4\pi r^2$ and the intensity is $I = \frac{P}{4\pi r^2}$. As the wave moves out from a source, the energy is conserved, but the intensity decreases as the area increases.

Conceptual Questions

Exercise:

Problem:

Consider a string with under tension with a constant linear mass density. A sinusoidal wave with an angular frequency and amplitude produced by some external driving force. If the frequency of the driving force is decreased to half of the original frequency, how is the time-averaged power of the wave affected? If the amplitude of the driving force is decreased by half, how is the time-averaged power affected? Explain your answer.

Solution:

The time averaged power is $P = \frac{E_\lambda}{T} = \frac{1}{2}\mu A^2\omega^2\frac{\lambda}{T} = \frac{1}{2}\mu A^2\omega^2 v$. If the frequency or amplitude is halved, the power decreases by a factor of 4.

Exercise:

Problem:

Circular water waves decrease in amplitude as they move away from where a rock is dropped. Explain why.

Exercise:

Problem:

In a transverse wave on a string, the motion of the string is perpendicular to the motion of the wave. If this is so, how is possible to move energy along the length of the string?

Solution:

As a portion on the string moves vertically, it exerts a force on the neighboring portion of the string, doing work on the portion and transferring the energy.

Exercise:

Problem:

The energy from the sun warms the portion of the earth facing the sun during the daylight hours. Why are the North and South Poles cold while the equator is quite warm?

Exercise:

Problem:

The intensity of a spherical waves decreases as the wave moves away from the source. If the intensity of the wave at the source is I_0, how far from the source will the intensity decrease by a factor of nine?

Solution:

The intensity of a spherical wave is $I = \frac{P}{4\pi r^2}$, if no energy is dissipated the intensity will decrease by a factor of nine at three meters.

Problems

Exercise:

Problem:

A string of length 5 m and a mass of 90 g is held under a tension of 100 N. A wave travels down the string that is modeled as $y\left(x,t\right)=0.01\,\text{m}\,\sin\left(15.7\,\text{m}^{-1}x-1170.12\,\text{s}^{-1}\right)$. What is the power over one wavelength?

Solution:

$v=74.54\,\text{m/s},\,P_{\lambda}=91.85\,\text{W}$

Exercise:

Problem:

Ultrasound of intensity $1.50\,\times\,10^{2}\,\text{W/m}^{2}$ is produced by the rectangular head of a medical imaging device measuring 3.00 cm by 5.00 cm. What is its power output?

Exercise:

Problem:

The low-frequency speaker of a stereo set has a surface area of $A=0.05\,\text{m}^{2}$ and produces 1 W of acoustical power. (a) What is the intensity at the speaker? (b) If the speaker projects sound uniformly in all directions, at what distance from the speaker is the intensity $0.1\,\text{W/m}^{2}$?

Solution:

a. $I=20.0\,\text{W/m}^{2}$; b. $\begin{array}{l}I=\frac{P}{A},\,A=10.0\,\text{m}^{2}\\ A=4\pi r^{2},\,r=0.892\,\text{m}\end{array}$

Exercise:

Problem:

To increase the intensity of a wave by a factor of 50, by what factor should the amplitude be increased?

Exercise:

Problem:

A device called an insolation meter is used to measure the intensity of sunlight. It has an area of $100\,\text{cm}^{2}$ and registers 6.50 W. What is the intensity in W/m^{2}?

Solution:

$I=650\,\text{W/m}^{2}$

Exercise:

Problem:

Energy from the Sun arrives at the top of Earth's atmosphere with an intensity of $1400\ \text{W/m}^2$. How long does it take for $1.80\ \times\ 10^9$ J to arrive on an area of $1.00\ \text{m}^2$?

Exercise:

Problem:

Suppose you have a device that extracts energy from ocean breakers in direct proportion to their intensity. If the device produces 10.0 kW of power on a day when the breakers are 1.20 m high, how much will it produce when they are 0.600 m high?

Solution:

$$P \propto E \propto I \propto X^2 \Rightarrow \frac{P_2}{P_1} = \left(\frac{X_2}{X_1}\right)^2$$

$$P_2 = 2.50\ \text{kW}$$

Exercise:

Problem:

A photovoltaic array of (solar cells) is 10.0% efficient in gathering solar energy and converting it to electricity. If the average intensity of sunlight on one day is $70.00\ \text{W/m}^2$, what area should your array have to gather energy at the rate of 100 W? (b) What is the maximum cost of the array if it must pay for itself in two years of operation averaging 10.0 hours per day? Assume that it earns money at the rate of 9.00 cents per kilowatt-hour.

Exercise:

Problem:

A microphone receiving a pure sound tone feeds an oscilloscope, producing a wave on its screen. If the sound intensity is originally $2.00\ \times\ 10^{-5}\ \text{W/m}^2$, but is turned up until the amplitude increases by 30.0%, what is the new intensity?

Solution:

$$I \propto X^2 \Rightarrow \frac{I_1}{I_2} = \left(\frac{X_1}{X_2}\right)^2 \Rightarrow$$

$$I_2 = 3.38\ \times\ 10^{-5}\ \text{W/m}^2$$

Exercise:

Problem:

A string with a mass of 0.30 kg has a length of 4.00 m. If the tension in the string is 50.00 N, and a sinusoidal wave with an amplitude of 2.00 cm is induced on the string, what must the frequency be for an average power of 100.00 W?

Exercise:

Problem:

The power versus time for a point on a string $(\mu = 0.05\,\text{kg/m})$ in which a sinusoidal traveling wave is induced is shown in the preceding figure. The wave is modeled with the wave equation $y\left(x,t\right) = A\,\text{sin}\left(20.93\,\text{m}^{-1}x - \omega t\right)$. What is the frequency and amplitude of the wave?

Solution:

$f = 100.00\,\text{Hz},\ A = 1.10\,\text{cm}$

Exercise:

Problem:

A string is under tension F_{T1}. Energy is transmitted by a wave on the string at rate P_1 by a wave of frequency f_1. What is the ratio of the new energy transmission rate P_2 to P_1 if the tension is doubled?

Exercise:

Problem:

A 250-Hz tuning fork is struck and the intensity at the source is I_1 at a distance of one meter from the source. (a) What is the intensity at a distance of 4.00 m from the source? (b) How far from the tuning fork is the intensity a tenth of the intensity at the source?

Solution:

a. $I_2 = 0.063 I_1$; b. $\begin{array}{rcl} I_1 4\pi r_1^2 & = & I_2 4\pi r_2^2 \\ r_2 & = & 3.16\,\text{m} \end{array}$

Exercise:

Problem:

A sound speaker is rated at a voltage of $P = 120.00\,\text{V}$ and a current of $I = 10.00\,\text{A}$. Electrical power consumption is $P = IV$. To test the speaker, a signal of a sine wave is applied to the speaker. Assuming that the sound wave moves as a spherical wave and that all of the energy applied to the speaker is converted to sound energy, how far from the speaker is the intensity equal to $3.82\,\text{W/m}^2$?

Exercise:

Problem:

The energy of a ripple on a pond is proportional to the amplitude squared. If the amplitude of the ripple is 0.1 cm at a distance from the source of 6.00 meters, what was the amplitude at a distance of 2.00 meters from the source?

Solution:

$$2\pi r_1 A_1^2 = 2\pi r_2 A_2^2,\ A_1 = \left(\frac{r_2}{r_1}\right)^{1/2} A_1 = 0.17\,\mathrm{m}$$

Glossary

intensity (*I*)
power per unit area

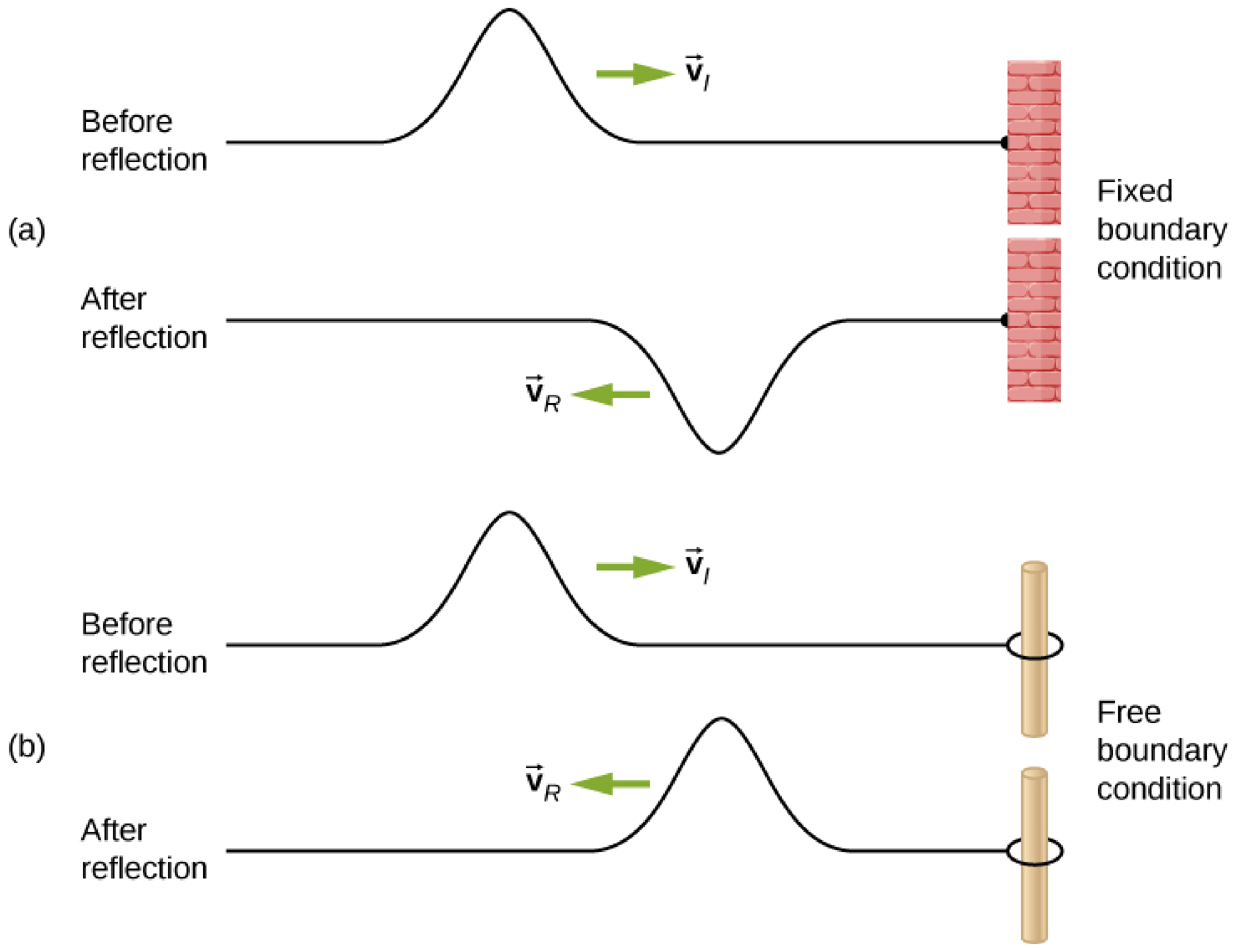
Interference of Waves
By the end of this section, you will be able to:

- Explain how mechanical waves are reflected and transmitted at the boundaries of a medium
- Define the terms interference and superposition
- Find the resultant wave of two identical sinusoidal waves that differ only by a phase shift

Up to now, we have been studying mechanical waves that propagate continuously through a medium, but we have not discussed what happens when waves encounter the boundary of the medium or what happens when a wave encounters another wave propagating through the same medium. Waves do interact with boundaries of the medium, and all or part of the wave can be reflected. For example, when you stand some distance from a rigid cliff face and yell, you can hear the sound waves reflect off the rigid surface as an echo. Waves can also interact with other waves propagating in the same medium. If you throw two rocks into a pond some distance from one another, the circular ripples that result from the two stones seem to pass through one another as they propagate out from where the stones entered the water. This phenomenon is known as interference. In this section, we examine what happens to waves encountering a boundary of a medium or another wave propagating in the same medium. We will see that their behavior is quite different from the behavior of particles and rigid bodies. Later, when we study modern physics, we will see that only at the scale of atoms do we see similarities in the properties of waves and particles.

Reflection and Transmission

When a wave propagates through a medium, it reflects when it encounters the boundary of the medium. The wave before hitting the boundary is known as the incident wave. The wave after encountering the boundary is known as the reflected wave. How the wave is reflected at the boundary of the medium depends on the boundary conditions; waves will react differently if the boundary of the medium is fixed in place or free to move ([link]). A **fixed boundary condition** exists when the medium at a boundary is fixed in place so it cannot move. A **free boundary condition** exists when the medium at the boundary is free to move.

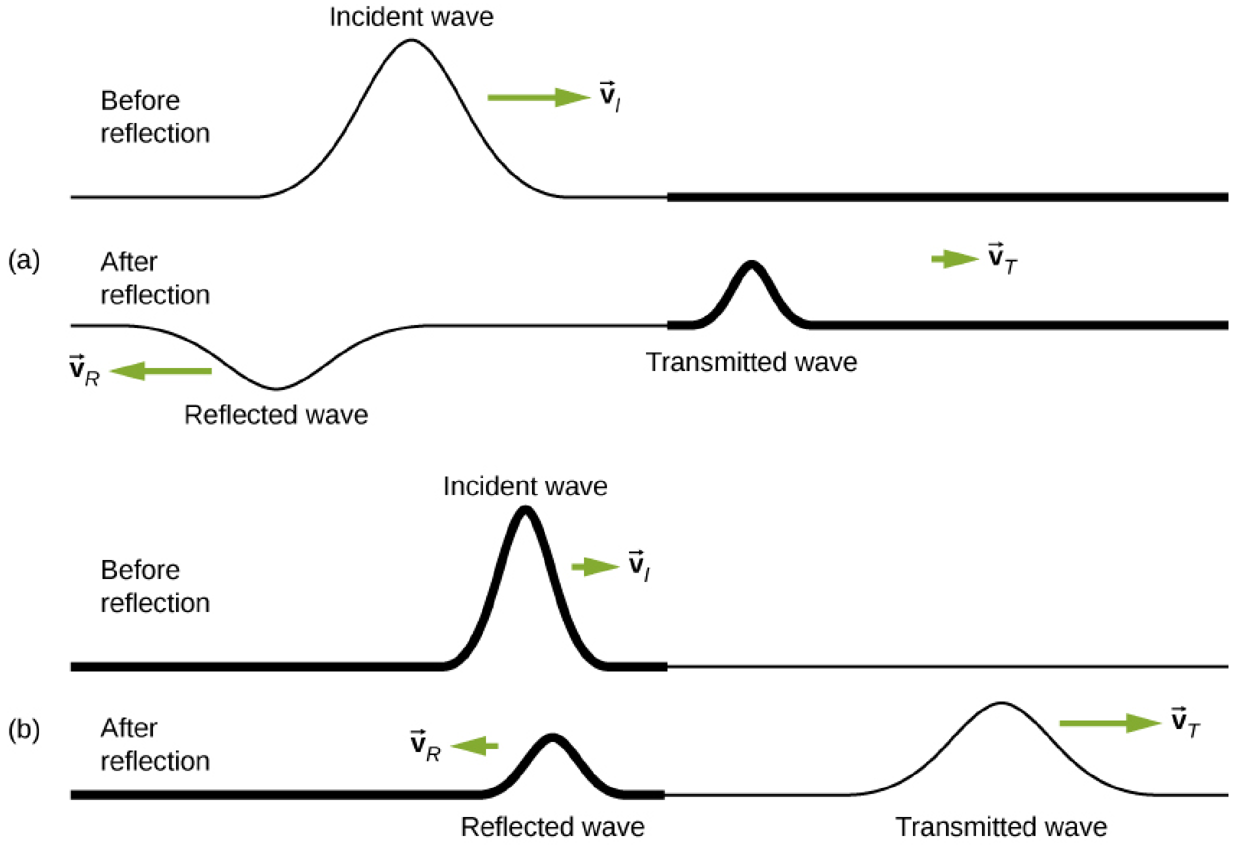


(a) One end of a string is fixed so that it cannot move. A wave propagating on the string, encountering this *fixed boundary condition*, is reflected $180^\circ\,(\pi\,\text{rad})$ out of phase with respect to the incident wave. (b) One end of a string is tied to a solid ring of negligible mass on a frictionless lab pole, where the ring is free to move. A wave propagating on the string, encountering this *free boundary condition*, is reflected in phase $0^\circ\,(0\,\text{rad})$ with respect to the wave.

Part (a) of the [link] shows a fixed boundary condition. Here, one end of the string is fixed to a wall so the end of the string is fixed in place and the medium (the string) at the boundary cannot move. When the wave is reflected, the amplitude of the reflected way is exactly the same as the amplitude of the incident wave, but the reflected wave is reflected $180^\circ\;(\pi\,\text{rad})$ out of phase with respect to the incident wave. The phase change can be explained using Newton’s third law: Recall that Newton’s third law states that when object *A* exerts a force on object *B*, then object *B* exerts an equal and opposite force on object *A*. As the incident wave encounters the wall, the string exerts an upward force on the wall and the wall reacts by exerting an equal and opposite force on the string. The reflection at a fixed boundary is inverted. Note that the figure shows a crest of the incident wave reflected as a trough. If the incident wave were a trough, the reflected wave would be a crest.

Part (b) of the figure shows a free boundary condition. Here, one end of the string is tied to a solid ring of negligible mass on a frictionless pole, so the end of the string is free to move up and down. As the incident wave encounters the boundary of the medium, it is also reflected. In the case of a free boundary condition, the reflected wave is in phase with respect to the incident wave. In this case, the wave encounters the free boundary applying an upward force on the ring, accelerating the ring up. The ring travels up to the maximum height equal to the amplitude of the wave and then accelerates down towards the equilibrium position due to the tension in the string. The figure shows the crest of an incident wave being reflected in phase with respect to the incident wave as a crest. If the incident wave were a trough, the reflected wave would also be a trough. The amplitude of the reflected wave would be equal to the amplitude of the incident wave.

In some situations, the boundary of the medium is neither fixed nor free. Consider [link](a), where a low-linear mass density string is attached to a string of a higher linear mass density. In this case, the reflected wave is out of phase with respect to the incident wave. There is also a transmitted wave that is in phase with respect to the incident wave. Both the transmitted and the reflected waves have amplitudes less than the amplitude of the incident wave. If the tension is the same in both strings, the wave speed is higher in the string with the lower linear mass density.



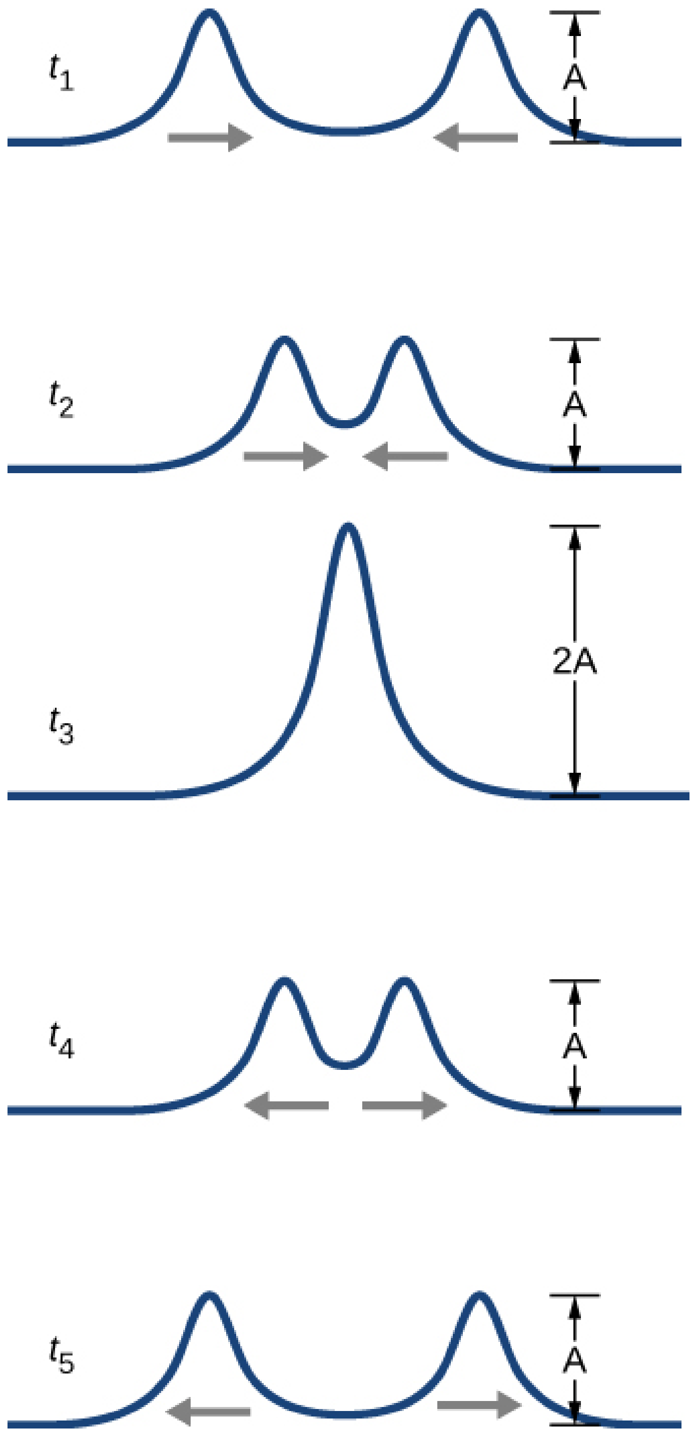
Waves traveling along two types of strings: a thick string with a high linear density and a thin string with a low linear density. Both strings are under the same tension, so a wave moves faster on the low-density string than on the high-density string. (a) A wave moving from a low-density to a high-density medium results in a reflected wave that is $180^\circ \, (\pi \, \text{rad})$ out of phase with respect to the incident pulse (or wave) and a transmitted wave that is in phase with the incident wave. (b) When a wave moves from a high-density medium to a low-density medium, both the reflected and transmitted wave are in phase with respect to the incident wave.

Part (b) of the figure shows a high-linear mass density string is attached to a string of a lower linear density. In this case, the reflected wave is in phase with respect to the incident wave. There is also a transmitted wave that is in phase with respect to the incident wave. Both the incident and the reflected waves have amplitudes less than the amplitude of the incident wave. Here you may notice that if the tension is the same in both strings, the wave speed is higher in the string with the lower linear mass density.

Superposition and Interference

Most waves do not look very simple. Complex waves are more interesting, even beautiful, but they look formidable. Most interesting mechanical waves consist of a combination of two or more traveling waves propagating in the same medium. The principle of superposition can be used to analyze the combination of waves.

Consider two simple pulses of the same amplitude moving toward one another in the same medium, as shown in [link]. Eventually, the waves overlap, producing a wave that has twice the amplitude, and then continue on unaffected by the encounter. The pulses are said to interfere, and this phenomenon is known as **interference**.



Two pulses moving toward one another experience interference. The term interference refers to what happens when two waves overlap.

To analyze the interference of two or more waves, we use the principle of superposition. For mechanical waves, the principle of **superposition** states that if two or more traveling waves combine at the same point, the resulting position of the mass element of the medium, at that point, is the algebraic sum of the position due to the individual waves. This property is exhibited by many waves observed, such as waves on a string, sound waves, and surface water waves. Electromagnetic waves also obey the superposition principle, but the electric and magnetic fields of the combined wave are added instead of the displacement of the medium. Waves that obey the superposition principle are linear waves; waves that do not obey the superposition principle are said to be nonlinear waves. In this chapter, we deal with linear waves, in particular, sinusoidal waves.

The superposition principle can be understood by considering the linear wave equation. In Mathematics of a Wave, we defined a linear wave as a wave whose mathematical representation obeys the linear wave equation. For a transverse wave on a string with an elastic restoring force, the linear wave equation is

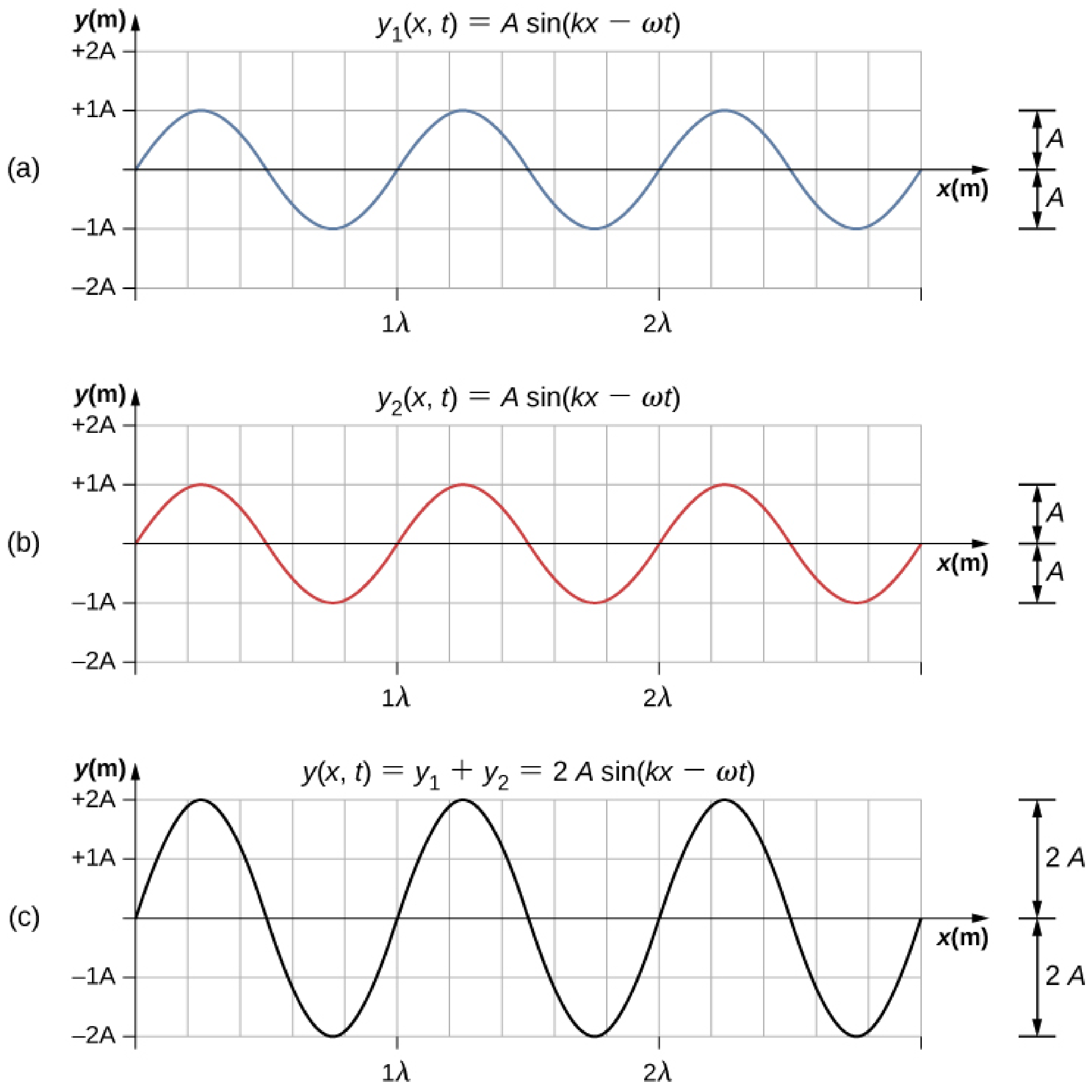
Equation:

$$\frac{\partial^2 y\left(x,t\right)}{\partial x^2} = \frac{1}{v^2}\,\frac{\partial^2 y\left(x,t\right)}{\partial t^2}.$$

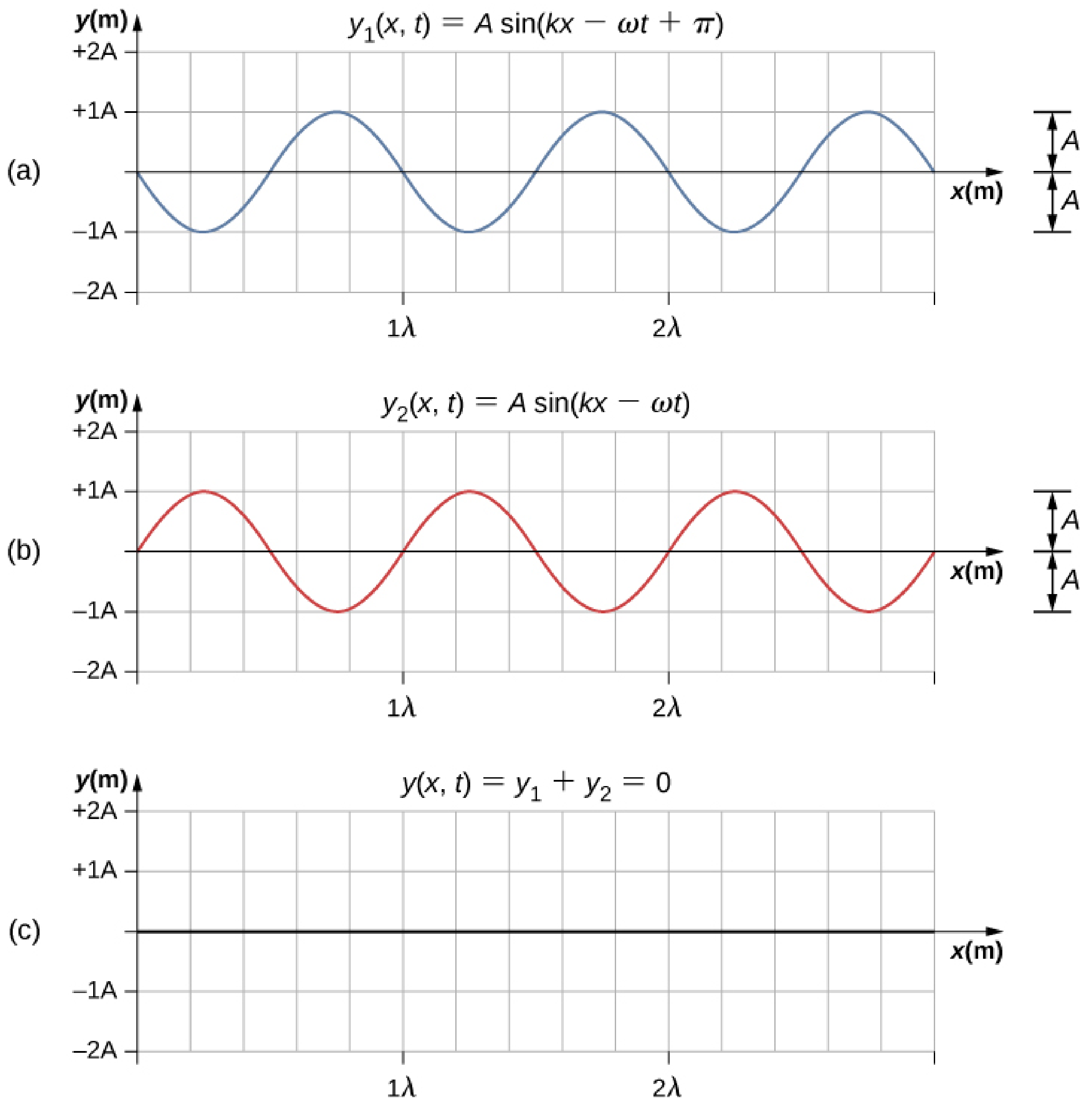
Any wave function $y(x,t) = y\left(x \mp vt\right)$, where the argument of the function is linear $\left(x \mp vt\right)$ is a solution to the linear wave equation and is a linear wave function. If wave functions $y_1\left(x,t\right)$ and $y_2\left(x,t\right)$ are solutions to the linear wave equation, the sum of the two functions $y_1\left(x,t\right) + y_2\left(x,t\right)$ is also a solution to the linear wave equation. Mechanical waves that obey superposition are normally restricted to waves with amplitudes that are small with respect to their wavelengths. If the amplitude is too large, the medium is distorted past the region where the restoring force of the medium is linear.

Waves can interfere constructively or destructively. [link] shows two identical sinusoidal waves that arrive at the same point exactly in phase. [link](a) and (b) show the two individual waves, [link](c) shows the resultant wave that results from the algebraic sum of the two linear waves. The crests of the two waves are precisely aligned, as are the troughs. This superposition produces **constructive interference**. Because the disturbances add, constructive interference produces a wave that has twice the amplitude of the individual waves, but has the same wavelength.

[link] shows two identical waves that arrive exactly 180° out of phase, producing **destructive interference**. [link](a) and (b) show the individual waves, and [link](c) shows the superposition of the two waves. Because the troughs of one wave add the crest of the other wave, the resulting amplitude is zero for destructive interference—the waves completely cancel.

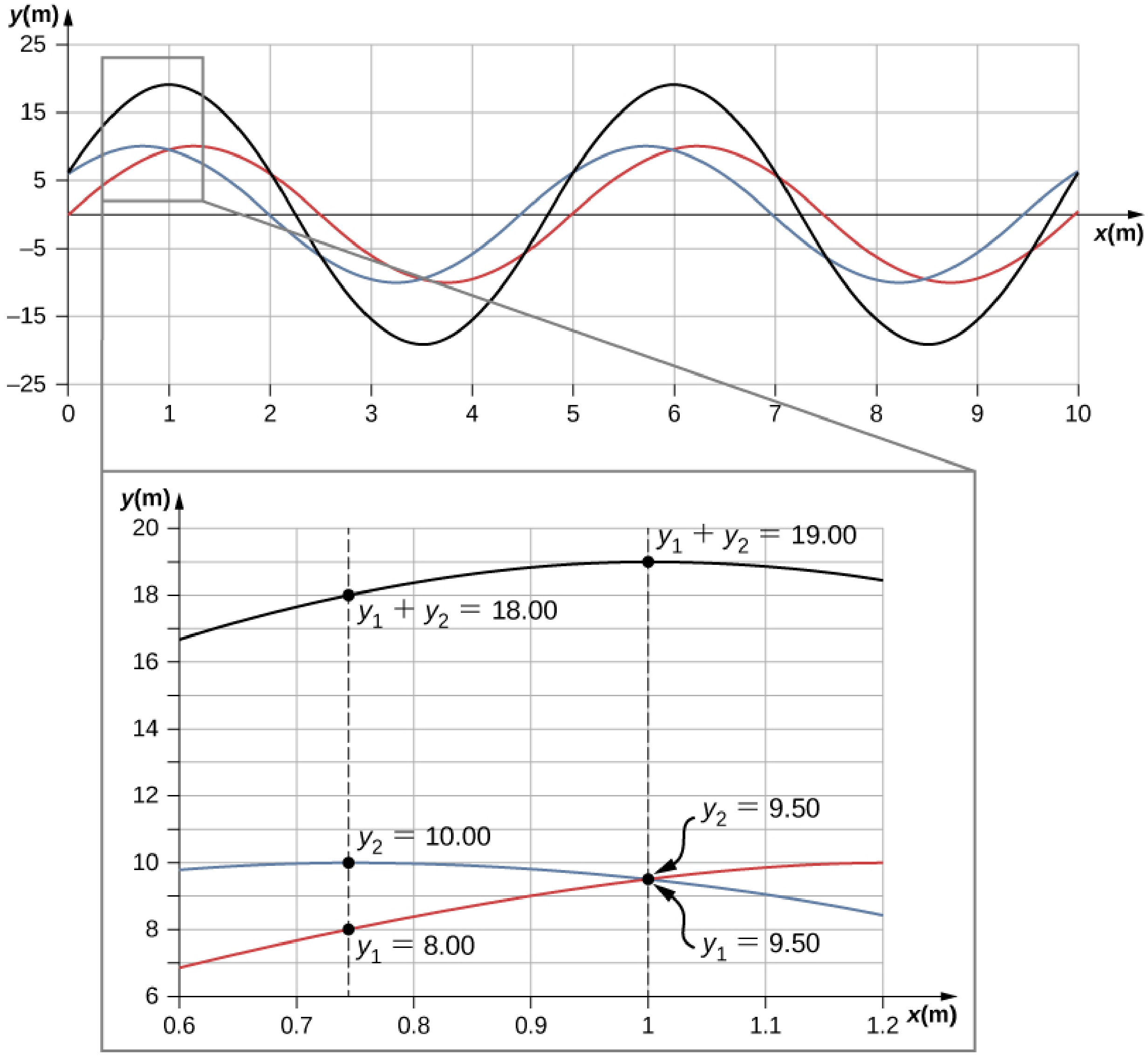


Constructive interference of two identical waves produces a wave with twice the amplitude, but the same wavelength.



Destructive interference of two identical waves, one with a phase shift of 180° (π rad), produces zero amplitude, or complete cancellation.

When linear waves interfere, the resultant wave is just the algebraic sum of the individual waves as stated in the principle of superposition. [link] shows two waves (red and blue) and the resultant wave (black). The resultant wave is the algebraic sum of the two individual waves.

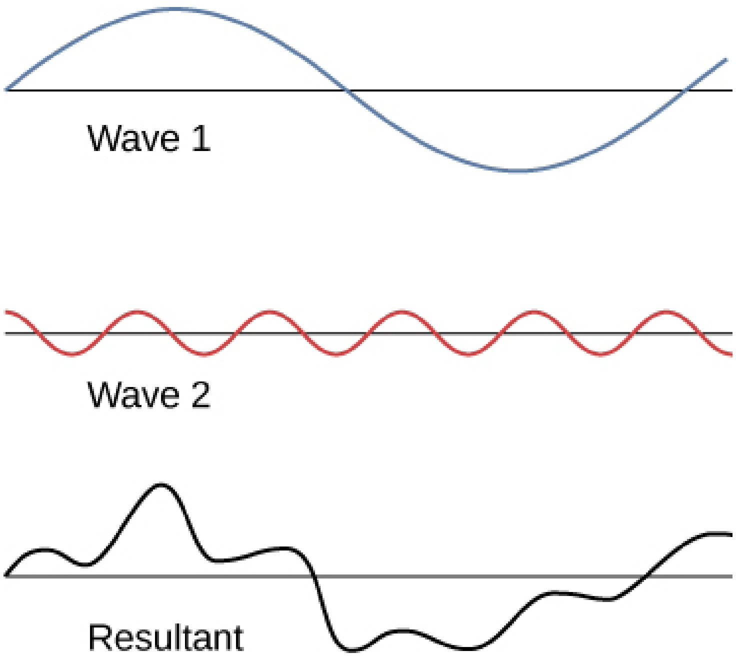


When two linear waves in the same medium interfere, the height of resulting wave is the sum of the heights of the individual waves, taken point by point. This plot shows two waves (red and blue) added together, along with the resulting wave (black). These graphs represent the height of the wave at each point. The waves may be any linear wave, including ripples on a pond, disturbances on a string, sound, or electromagnetic waves.

The superposition of most waves produces a combination of constructive and destructive interference, and can vary from place to place and time to time.

Sound from a stereo, for example, can be loud in one spot and quiet in another. Varying loudness means the sound waves add partially constructively and partially destructively at different locations. A stereo has at least two speakers creating sound waves, and waves can reflect from walls. All these waves interfere, and the resulting wave is the superposition of the waves.

We have shown several examples of the superposition of waves that are similar. [link] illustrates an example of the superposition of two dissimilar waves. Here again, the disturbances add, producing a resultant wave.



Superposition of nonidentical waves exhibits both constructive and destructive interference.

At times, when two or more mechanical waves interfere, the pattern produced by the resulting wave can be rich in complexity, some without any readily discernable patterns. For example, plotting the sound wave of your favorite music can look quite complex and is the superposition of the individual sound waves from many instruments; it is the complexity that makes the music interesting and worth listening to. At other times, waves can interfere and produce interesting phenomena, which are complex in their appearance and yet beautiful in simplicity of the physical principle of superposition, which formed the resulting wave. One example is the phenomenon known as standing waves, produced by two identical waves moving in different directions. We will look more closely at this phenomenon in the next section.

Note:
Try this simulation to make waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern. You can observe one source or two sources. Using two sources, you can observe the interference patterns that result from varying the frequencies and the amplitudes of the sources.

Superposition of Sinusoidal Waves that Differ by a Phase Shift

Many examples in physics consist of two sinusoidal waves that are identical in amplitude, wave number, and angular frequency, but differ by a phase shift:

Equation:

$$\begin{aligned} y_1(x,t) &= A\sin(kx-\omega t+\phi), \\ y_2(x,t) &= A\sin(kx-\omega t). \end{aligned}$$

When these two waves exist in the same medium, the resultant wave resulting from the superposition of the two individual waves is the sum of the two individual waves:

Equation:

$$y_R(x,t) = y_1(x,t) + y_2(x,t) = A\sin(kx-\omega t+\phi) + A\sin(kx-\omega t).$$

The resultant wave can be better understood by using the trigonometric identity:

Equation:

$$\sin u + \sin v = 2 \sin \left(\frac{u+v}{2} \right) \cos \left(\frac{u-v}{2} \right),$$

where $u = kx - \omega t + \phi$ and $v = kx - \omega t$. The resulting wave becomes

Equation:

$$\begin{aligned} y_R(x,t) &= y_1(x,t) + y_2(x,t) = A \sin(kx - \omega t + \phi) + A \sin(kx - \omega t) \\ &= 2A \sin\left(\tfrac{(kx-\omega t+\phi)+(kx-\omega t)}{2}\right) \cos\left(\tfrac{(kx-\omega t+\phi)-(kx-\omega t)}{2}\right) \\ &= 2A \sin\left(kx - \omega t + \tfrac{\phi}{2}\right) \cos\left(\tfrac{\phi}{2}\right). \end{aligned}$$

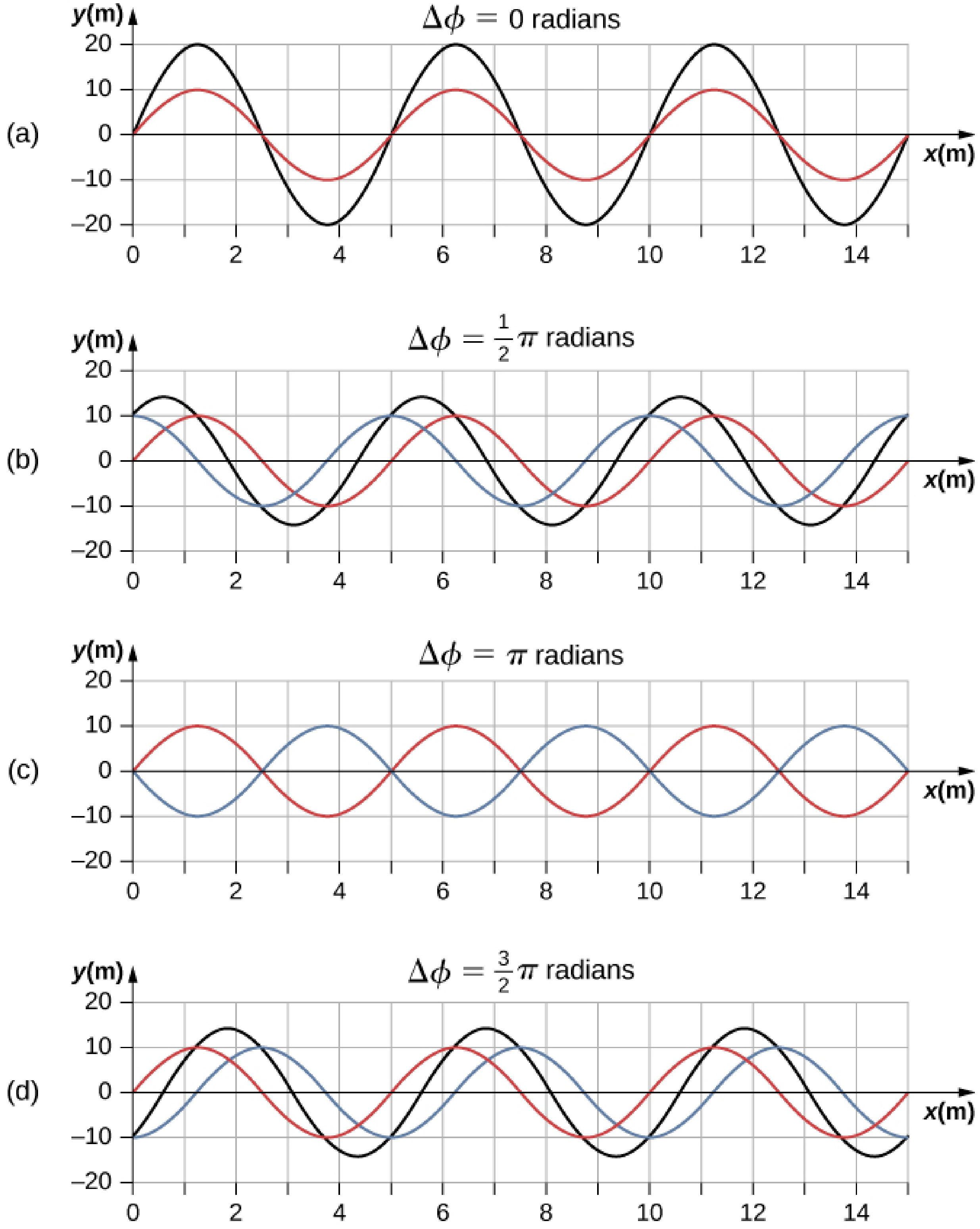
This equation is usually written as

Note:

Equation:

$$y_R(x,t) = \left[2A \cos\left(\frac{\phi}{2}\right)\right] \sin\left(kx - \omega t + \frac{\phi}{2}\right).$$

The resultant wave has the same wave number and angular frequency, an amplitude of $A_R = \left[2A \cos\left(\frac{\phi}{2}\right)\right]$, and a phase shift equal to half the original phase shift. Examples of waves that differ only in a phase shift are shown in [link]. The red and blue waves each have the same amplitude, wave number, and angular frequency, and differ only in a phase shift. They therefore have the same period, wavelength, and frequency. The green wave is the result of the superposition of the two waves. When the two waves have a phase difference of zero, the waves are in phase, and the resultant wave has the same wave number and angular frequency, and an amplitude equal to twice the individual amplitudes (part (a)). This is constructive interference. If the phase difference is 180°, the waves interfere in destructive interference (part (c)). The resultant wave has an amplitude of zero. Any other phase difference results in a wave with the same wave number and angular frequency as the two incident waves but with a phase shift of $\phi/2$ and an amplitude equal to $2A\cos(\phi/2)$. Examples are shown in parts (b) and (d).



Superposition of two waves with identical amplitudes, wavelengths, and frequency, but that differ in a phase shift. The red wave is defined by the wave function $y_1(x,t) = A\sin(kx - \omega t)$ and the blue wave is defined by the wave function $y_2(x,t) = A\sin(kx - \omega t + \phi)$. The black line shows the result of adding the two waves. The phase difference between the two waves are (a) $0.00\,\text{rad}$, (b) $\pi/2\,\text{rad}$, (c) $\pi\,\text{rad}$, and (d) $3\pi/2\,\text{rad}$.

Summary

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs from the superposition of two identical waves that are in phase.
- Destructive interference occurs from the superposition of two identical waves that are 180° (π radians) out of phase.
- The wave that results from the superposition of two sine waves that differ only by a phase shift is a wave with an amplitude that depends on the value of the phase difference.

Conceptual Questions

Exercise:

Problem:

An incident sinusoidal wave is sent along a string that is fixed to the wall with a wave speed of *v*. The wave reflects off the end of the string. Describe the reflected wave.

Exercise:

Problem:

A string of a length of 2.00 m with a linear mass density of $\mu = 0.006\,\mathrm{kg/m}$ is attached to the end of a 2.00-m-long string with a linear mass density of $\mu = 0.012\,\mathrm{kg/m}$. The free end of the higher-density string is fixed to the wall, and a student holds the free end of the low-density string, keeping the tension constant in both strings. The student sends a pulse down the string. Describe what happens at the interface between the two strings.

Solution:

At the interface, the incident pulse produces a reflected pulse and a transmitted pulse. The reflected pulse would be out of phase with respect to the incident pulse, and would move at the same propagation speed as the incident pulse, but would move in the opposite direction. The transmitted pulse would travel in the same direction as the incident pulse, but at half the speed. The transmitted pulse would be in phase with the incident pulse. Both the reflected pulse and the transmitted pulse would have amplitudes less than the amplitude of the incident pulse.

Exercise:

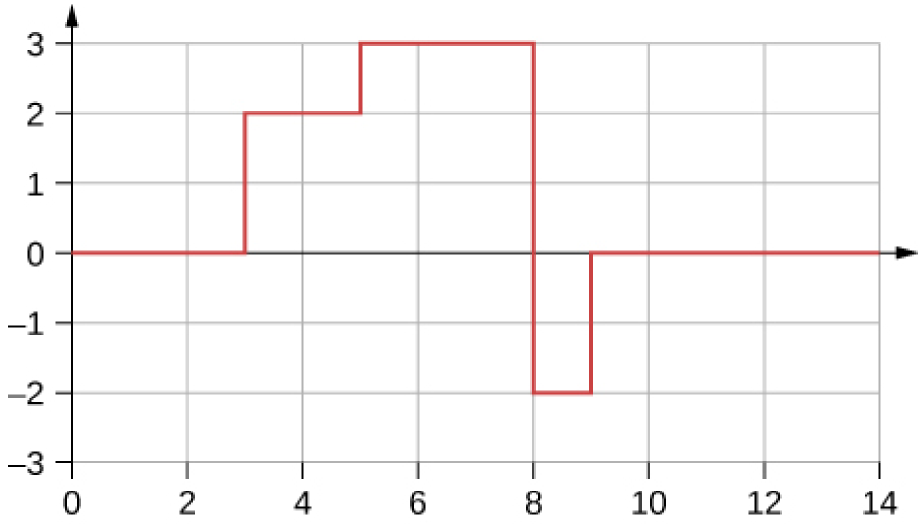
Problem:

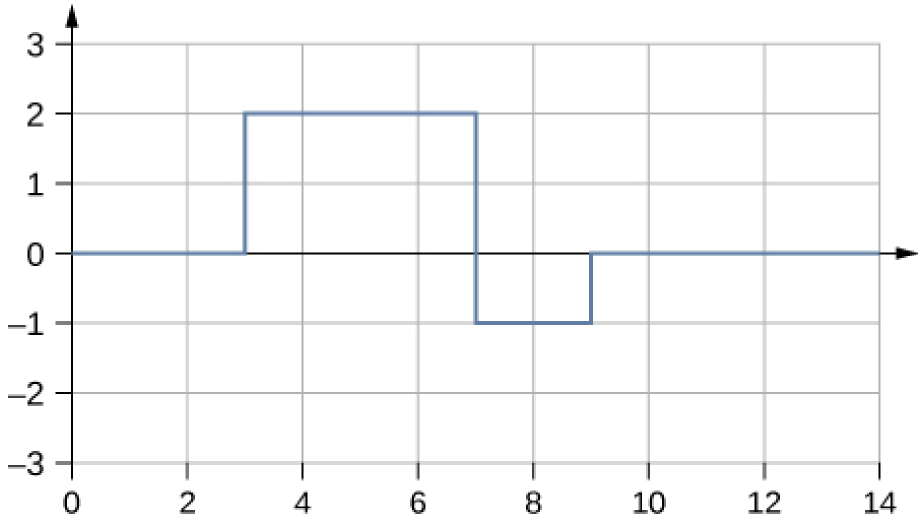
A long, tight spring is held by two students, one student holding each end. Each student gives the end a flip sending one wavelength of a sinusoidal wave down the spring in opposite directions. When the waves meet in the middle, what does the wave look like?

Exercise:

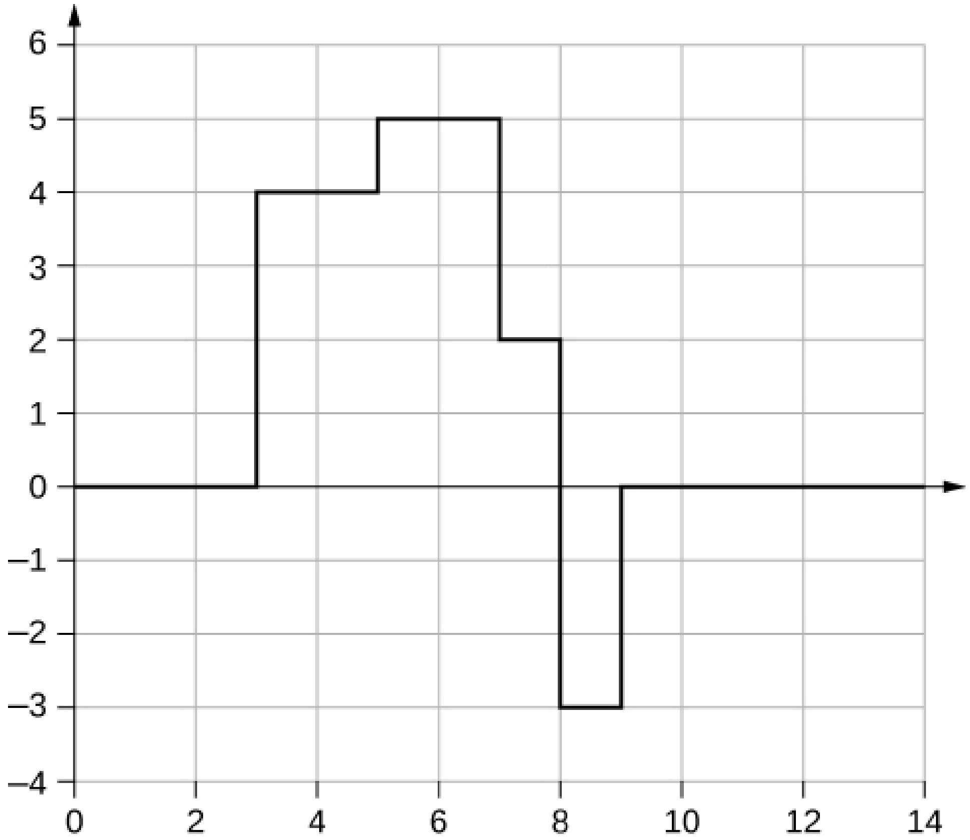
Problem:

Many of the topics discussed in this chapter are useful beyond the topics of mechanical waves. It is hard to conceive of a mechanical wave with sharp corners, but you could encounter such a wave form in your digital electronics class, as shown below. This could be a signal from a device known as an analog to digital converter, in which a continuous voltage signal is converted into a discrete signal or a digital recording of sound. What is the result of the superposition of the two signals?





Solution:



Exercise:

Problem:

A string of a constant linear mass density is held taut by two students, each holding one end. The tension in the string is constant. The students each send waves down the string by wiggling the string. (a) Is it possible for the waves to have different wave speeds? (b) Is it possible for the waves to have different frequencies? (c) Is it possible for the waves to have different wavelengths?

Problems

Exercise:

Problem:

Consider two sinusoidal waves traveling along a string, modeled as $y_1(x,t) = 0.3\,\text{m}\sin\left(4\,\text{m}^{-1}x + 3\,\text{s}^{-1}t\right)$ and $y_2(x,t) = 0.6\,\text{m}\sin\left(8\,\text{m}^{-1}x - 6\,\text{s}^{-1}t\right)$. What is the height of the resultant wave formed by the interference of the two waves at the position $x = 0.5\,\text{m}$ at time $t = 0.2\,\text{s}$?

Exercise:

Problem:

Consider two sinusoidal sine waves traveling along a string, modeled as $y_1(x,t) = 0.3\,\text{m}\sin\left(4\,\text{m}^{-1}x + 3\,\text{s}^{-1}t + \frac{\pi}{3}\right)$ and $y_2(x,t) = 0.6\,\text{m}\sin\left(8\,\text{m}^{-1}x - 6\,\text{s}^{-1}t\right)$. What is the height of the resultant wave formed by the interference of the two waves at the position $x = 1.0\,\text{m}$ at time $t = 3.0\,\text{s}$?

Solution:

$y(x,t) = 0.63\,\text{m}$

Exercise:

Problem:

Consider two sinusoidal sine waves traveling along a string, modeled as $y_1(x,t) = 0.3\,\text{m}\sin\left(4\,\text{m}^{-1}x - 3\,\text{s}^{-1}t\right)$ and $y_2(x,t) = 0.3\,\text{m}\sin\left(4\,\text{m}^{-1}x + 3\,\text{s}^{-1}t\right)$. What is the wave function of the resulting wave? [*Hint:* Use the trig identity $\sin(u \pm v) = \sin u \cos v \pm \cos u \sin v$

Exercise:

Problem:

Two sinusoidal waves are moving through a medium in the same direction, both having amplitudes of 3.00 cm, a wavelength of 5.20 m, and a period of 6.52 s, but one has a phase shift of an angle ϕ. What is the phase shift if the resultant wave has an amplitude of 5.00 cm? [*Hint:* Use the trig identity $\sin u + \sin v = 2\sin\left(\frac{u+v}{2}\right)\cos\left(\frac{u-v}{2}\right)$

Solution:

$A_R = 2A\cos\left(\frac{\phi}{2}\right),\ \phi = 1.17\,\text{rad}$

Exercise:

Problem:

Two sinusoidal waves are moving through a medium in the positive *x*-direction, both having amplitudes of 6.00 cm, a wavelength of 4.3 m, and a period of 6.00 s, but one has a phase shift of an angle $\phi = 0.50\,\text{rad}$. What is the height of the resultant wave at a time $t = 3.15\,\text{s}$ and a position $x = 0.45\,\text{m}$?

Exercise:

Problem:

Two sinusoidal waves are moving through a medium in the positive *x*-direction, both having amplitudes of 7.00 cm, a wave number of $k = 3.00\,\text{m}^{-1}$, an angular frequency of $\omega = 2.50\,\text{s}^{-1}$, and a period of 6.00 s, but one has a phase shift of an angle $\phi = \frac{\pi}{12}$ rad. What is the height of the resultant wave at a time $t = 2.00\,\text{s}$ and a position $x = 0.53\,\text{m}$?

Solution:

$y_R = 1.90\,\text{cm}$

Exercise:

Problem:

Consider two waves $y_1(x,t)$ and $y_2(x,t)$ that are identical except for a phase shift propagating in the same medium. (a)What is the phase shift, in radians, if the amplitude of the resulting wave is 1.75 times the amplitude of the individual waves? (b) What is the phase shift in degrees? (c) What is the phase shift as a percentage of the individual wavelength?

Exercise:

Problem:

Two sinusoidal waves, which are identical except for a phase shift, travel along in the same direction. The wave equation of the resultant wave is $y_R(x,t) = 0.70\,\text{m}\sin\left(3.00\,\text{m}^{-1}x - 6.28\,\text{s}^{-1}t + \pi/16\,\text{rad}\right)$. What are the angular frequency, wave number, amplitude, and phase shift of the individual waves?

Solution:

$$\begin{aligned} \omega &= 6.28\,\text{s}^{-1},\ k = 3.00\,\text{m}^{-1},\ \phi = \tfrac{\pi}{8}\,\text{rad}, \\ A_R &= 2A\cos\left(\tfrac{\phi}{2}\right),\ A = 0.37\,\text{m} \end{aligned}$$

Exercise:

Problem:

Two sinusoidal waves, which are identical except for a phase shift, travel along in the same direction. The wave equation of the resultant wave is $y_R(x,t) = 0.35\,\text{cm}\sin\left(6.28\,\text{m}^{-1}x - 1.57\,\text{s}^{-1}t + \tfrac{\pi}{4}\right)$. What are the period, wavelength, amplitude, and phase shift of the individual waves?

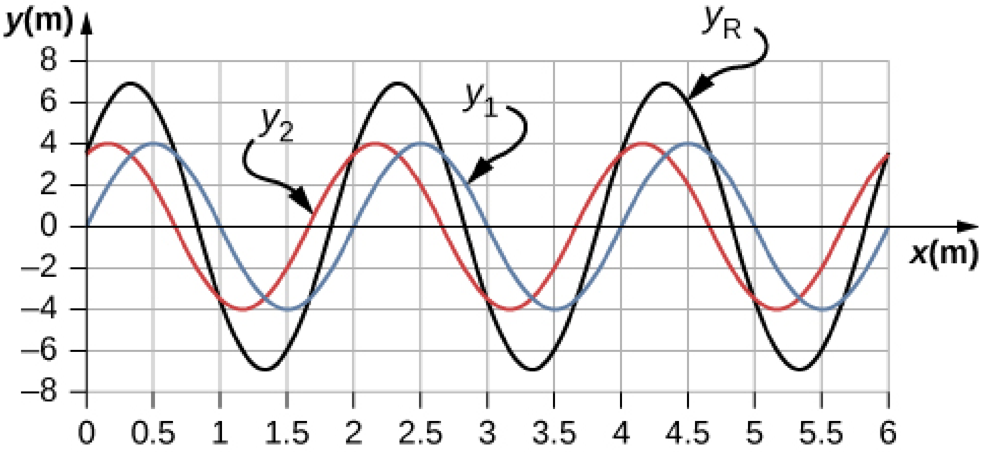
Exercise:

Problem:

Consider two wave functions, $y_1(x,t) = 4.00\,\text{m}\,\sin\left(\pi\,\text{m}^{-1}x - \pi\,\text{s}^{-1}t\right)$ and $y_2(x,t) = 4.00\,\text{m}\,\sin\left(\pi\,\text{m}^{-1}x - \pi\,\text{s}^{-1}t + \frac{\pi}{3}\right)$. (a) Using a spreadsheet, plot the two wave functions and the wave that results from the superposition of the two wave functions as a function of position $(0.00 \le x \le 6.00\,\text{m})$ for the time $t = 0.00\,\text{s}$. (b) What are the wavelength and amplitude of the two original waves? (c) What are the wavelength and amplitude of the resulting wave?

Solution:

a.



;
b. $\lambda = 2.0\,\text{m}$, $A = 4\,\text{m}$; c. $\lambda_R = 2.0\,\text{m}$, $A_R = 6.93\,\text{m}$

Exercise:

Problem:

Consider two wave functions, $y_2\left(x,t\right)=2.00\,\text{m}\sin\left(\frac{\pi}{2}\text{m}^{-1}x-\frac{\pi}{3}\text{s}^{-1}t\right)$ and $y_2\left(x,t\right)=2.00\,\text{m}\sin\left(\frac{\pi}{2}\text{m}^{-1}x-\frac{\pi}{3}\text{s}^{-1}t+\frac{\pi}{6}\right)$. (a) Verify that $y_R=2A\cos\left(\frac{\phi}{2}\right)\sin\left(kx-\omega t+\frac{\phi}{2}\right)$ is the solution for the wave that results from a superposition of the two waves. Make a column for *x*, y_1, y_2, y_1+y_2, and $y_R=2A\cos\left(\frac{\phi}{2}\right)\sin\left(kx-\omega t+\frac{\phi}{2}\right)$. Plot four waves as a function of position where the range of *x* is from 0 to 12 m.

Exercise:

Problem:

Consider two wave functions that differ only by a phase shift, $y_1\left(x,t\right)=A\cos\left(kx-\omega t\right)$ and $y_2\left(x,t\right)=A\cos\left(kx-\omega t+\phi\right)$. Use the trigonometric identities $\cos u+\cos v=2\cos\left(\frac{u-v}{2}\right)\cos\left(\frac{u+v}{2}\right)$ and $\cos\left(-\theta\right)=\cos\left(\theta\right)$ to find a wave equation for the wave resulting from the superposition of the two waves. Does the resulting wave function come as a surprise to you?

Solution:

$y_R\left(x,t\right)=2A\cos\left(\frac{\phi}{2}\right)\cos\left(kx-\omega t+\frac{\phi}{2}\right)$; The result is not surprising because $\cos\left(\theta\right)=\sin\left(\theta+\frac{\pi}{2}\right)$.

Glossary

constructive interference
: when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs

destructive interference
: when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough

fixed boundary condition
when the medium at a boundary is fixed in place so it cannot move

free boundary condition
exists when the medium at the boundary is free to move

interference
overlap of two or more waves at the same point and time

superposition
phenomenon that occurs when two or more waves arrive at the same point

Standing Waves and Resonance

By the end of this section, you will be able to:

- Describe standing waves and explain how they are produced
- Describe the modes of a standing wave on a string
- Provide examples of standing waves beyond the waves on a string

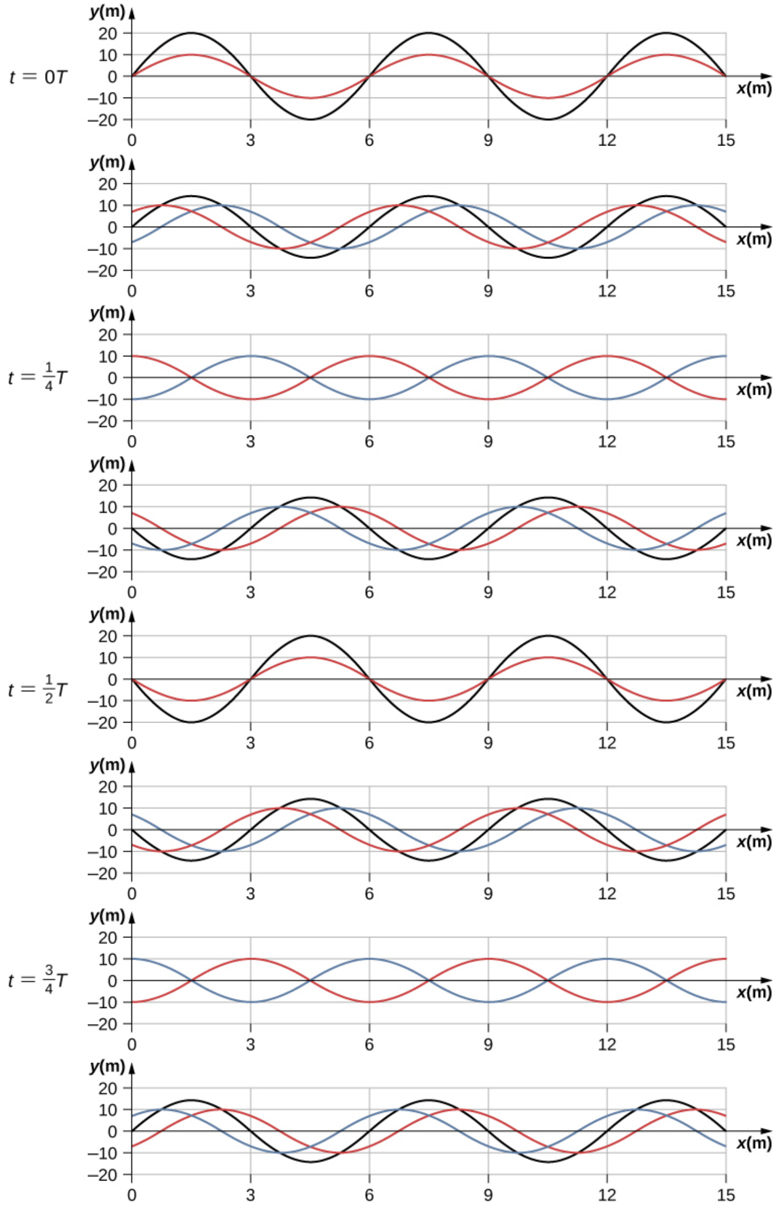
Throughout this chapter, we have been studying traveling waves, or waves that transport energy from one place to another. Under certain conditions, waves can bounce back and forth through a particular region, effectively becoming stationary. These are called **standing waves**.

Another related effect is known as resonance. In Oscillations, we defined resonance as a phenomenon in which a small-amplitude driving force could produce large-amplitude motion. Think of a child on a swing, which can be modeled as a physical pendulum. Relatively small-amplitude pushes by a parent can produce large-amplitude swings. Sometimes this resonance is good—for example, when producing music with a stringed instrument. At other times, the effects can be devastating, such as the collapse of a building during an earthquake. In the case of standing waves, the relatively large amplitude standing waves are produced by the superposition of smaller amplitude component waves.

Standing Waves

Sometimes waves do not seem to move; rather, they just vibrate in place. You can see unmoving waves on the surface of a glass of milk in a refrigerator, for example. Vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. [link] shows an experiment you can try at home. Take a bowl of milk and place it on a common box fan. Vibrations from the fan will produce circular standing waves in the milk. The waves are visible in the photo due to the reflection from a lamp. These waves are formed by the superposition of two or more traveling waves, such as illustrated in [link] for two identical waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. The resultant looks like a wave standing in place and, thus, is called a standing wave.

Standing waves are formed on the surface of a bowl of milk sitting on a box fan. The vibrations from the fan causes the surface of the milk to oscillate. The waves are visible due to the reflection of light from a lamp. (credit: David Chelton)



Time snapshots of two sine waves. The red wave is moving in the −*x*-direction and the blue wave is moving in the +*x*-direction. The resulting wave is shown in black. Consider the resultant wave at the points $x = 0\ \text{m}, 3\ \text{m}, 6\ \text{m}, 9\ \text{m}, 12\ \text{m}, 15\ \text{m}$ and notice that the resultant wave always equals zero at these points, no matter what the time is. These points are known as fixed points (nodes). In between each two nodes is an antinode, a place where the medium oscillates with an amplitude equal to the sum of the amplitudes of the individual waves.

Consider two identical waves that move in opposite directions. The first wave has a wave function of $y_1\left(x,t\right)=A\,\text{sin}\left(kx-\omega t\right)$ and the second wave has a wave function $y_2\left(x,t\right)=A\,\text{sin}\left(kx+\omega t\right)$. The waves interfere and form a resultant wave

Equation:

$$\begin{aligned}y\left(x,t\right)&=y_1\left(x,t\right)+y_2\left(x,t\right),\\ y\left(x,t\right)&=A\,\text{sin}\left(kx-\omega t\right)+A\,\text{sin}\left(kx+\omega t\right).\end{aligned}$$

This can be simplified using the trigonometric identity

Equation:

$$\text{sin}\left(\alpha\pm\beta\right)=\text{sin}\,\alpha\,\text{cos}\,\beta\pm\text{cos}\,\alpha\,\text{sin}\,\beta,$$

where $\alpha=kx$ and $\beta=\omega t$, giving us

Equation:

$$y\left(x,t\right)=A[\text{sin}\left(kx\right)\text{cos}\left(\omega t\right)-\text{cos}\left(kx\right)\text{sin}\left(\omega t\right)+\text{sin}\left(kx\right)\text{cos}\left(\omega t\right)+\text{cos}\left(kx\right)\text{sin}\left(\omega t\right)],$$

which simplifies to

Note:

Equation:

$$y\left(x,t\right)=\left[2A\,\text{sin}\left(kx\right)\right]\text{cos}(\omega t).$$

Notice that the resultant wave is a sine wave that is a function only of position, multiplied by a cosine function that is a function only of time. Graphs of *y*(*x*,*t*) as a function of *x* for various times are shown in [link]. The red wave moves in the negative *x*-direction, the blue wave moves in the positive *x*-direction, and the black wave is the sum of the two waves. As the red and blue waves move through each other, they move in and out of constructive interference and destructive interference.

Initially, at time $t=0$, the two waves are in phase, and the result is a wave that is twice the amplitude of the individual waves. The waves are also in phase at the time $t=\frac{T}{2}$. In fact, the waves are in phase at any integer multiple of half of a period:

Equation:

$$t=n\frac{T}{2}\text{ where }n=0,1,2,3....\text{ (in phase)}.$$

At other times, the two waves are $180°\left(\pi\text{ radians}\right)$ out of phase, and the resulting wave is equal to zero. This happens at

Equation:

$$t=\frac{1}{4}T,\ \frac{3}{4}T,\ \frac{5}{4}T,\ldots,\ \frac{n}{4}T \text{ where } n=1,3,5\ldots.\ \text{(out of phase)}.$$

Notice that some *x*-positions of the resultant wave are always zero no matter what the phase relationship is. These positions are called **nodes**. Where do the nodes occur? Consider the solution to the sum of the two waves

Equation:

$$y\left(x,t\right)=\left[2A\sin\left(kx\right)\right]\cos\left(\omega t\right).$$

Finding the positions where the sine function equals zero provides the positions of the nodes.

Equation:

$$\begin{array}{rcl}\sin\left(kx\right) & = & 0\\ kx & = & 0,\pi,2\pi,3\pi,\ldots\\ \frac{2\pi}{\lambda}x & = & 0,\pi,2\pi,3\pi,\ldots\\ x & = & 0,\frac{\lambda}{2},\lambda,\frac{3\lambda}{2},\ldots=n\frac{\lambda}{2}\quad n=0,1,2,3,\ldots.\end{array}$$

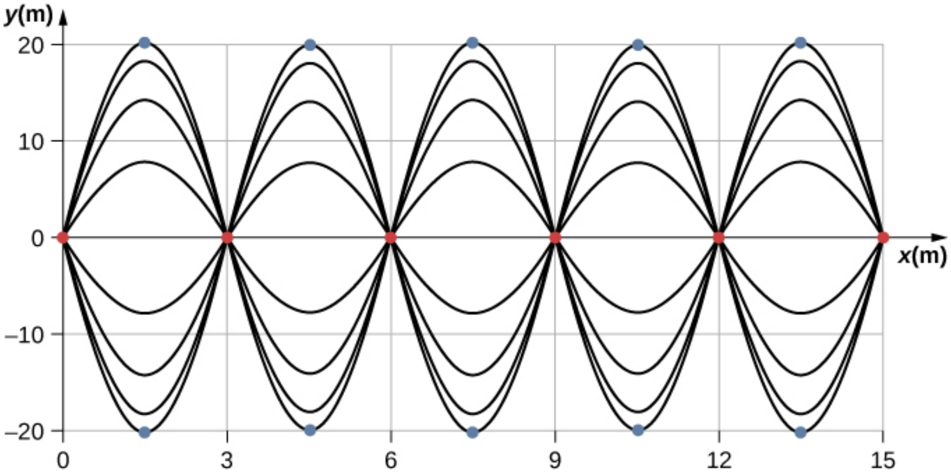
There are also positions where *y* oscillates between $y=\pm A$. These are the **antinodes**. We can find them by considering which values of *x* result in $\sin\left(kx\right)=\pm 1$.

Equation:

$$\begin{array}{rcl}\sin\left(kx\right) & = & \pm 1\\ kx & = & \frac{\pi}{2},\frac{3\pi}{2},\frac{5\pi}{2},\ldots\\ \frac{2\pi}{\lambda}x & = & \frac{\pi}{2},\frac{3\pi}{2},\frac{5\pi}{2},\ldots\\ x & = & \frac{\lambda}{4},\frac{3\lambda}{4},\frac{5\lambda}{4},\ldots=n\frac{\lambda}{4}\quad n=1,3,5,\ldots.\end{array}$$

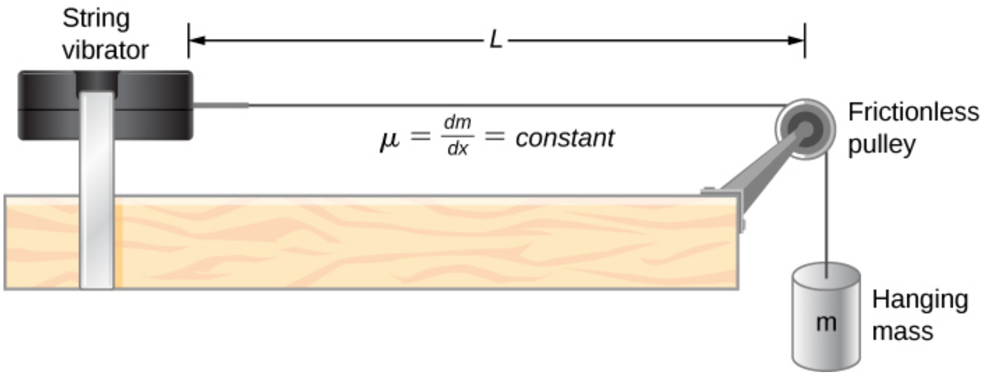
What results is a standing wave as shown in [link], which shows snapshots of the resulting wave of two identical waves moving in opposite directions. The resulting wave appears to be a sine wave with nodes at integer multiples of half wavelengths. The antinodes oscillate between $y=\pm 2A$ due to the cosine term, $\cos\left(\omega t\right)$, which oscillates between ± 1.

The resultant wave appears to be standing still, with no apparent movement in the *x*-direction, although it is composed of one wave function moving in the positive, whereas the second wave is moving in the negative *x*-direction. [link] shows various snapshots of the resulting wave. The nodes are marked with red dots while the antinodes are marked with blue dots.



When two identical waves are moving in opposite directions, the resultant wave is a standing wave. Nodes appear at integer multiples of half wavelengths. Antinodes appear at odd multiples of quarter wavelengths, where they oscillate between $y = \pm A$. The nodes are marked with red dots and the antinodes are marked with blue dots.

A common example of standing waves are the waves produced by stringed musical instruments. When the string is plucked, pulses travel along the string in opposite directions. The ends of the strings are fixed in place, so nodes appear at the ends of the strings—the boundary conditions of the system, regulating the resonant frequencies in the strings. The resonance produced on a string instrument can be modeled in a physics lab using the apparatus shown in [link].



A lab setup for creating standing waves on a string. The string has a node on each end and a constant linear density. The length between the fixed boundary conditions is L. The hanging mass provides the tension in the string, and the speed of the waves on the string is proportional to the square root of the tension divided by the linear mass density.

The lab setup shows a string attached to a string vibrator, which oscillates the string with an adjustable frequency *f*. The other end of the string passes over a frictionless pulley and is tied to a hanging mass. The magnitude of the tension in the string is equal to the weight of the hanging mass. The string has a constant linear density (mass per length) μ and the speed at which a wave travels down the string equals $v=\sqrt{\frac{F_T}{\mu}}=\sqrt{\frac{mg}{\mu}}$ [link]. The symmetrical boundary conditions (a node at each end) dictate the possible frequencies that can excite standing waves. Starting from a frequency of zero and slowly increasing the frequency, the first mode $n=1$ appears as shown in [link]. The first mode, also called the fundamental mode or the first harmonic, shows half of a wavelength has formed, so the wavelength is equal to twice the length between the nodes $\lambda_1=2L$. The **fundamental frequency**, or first harmonic frequency, that drives this mode is

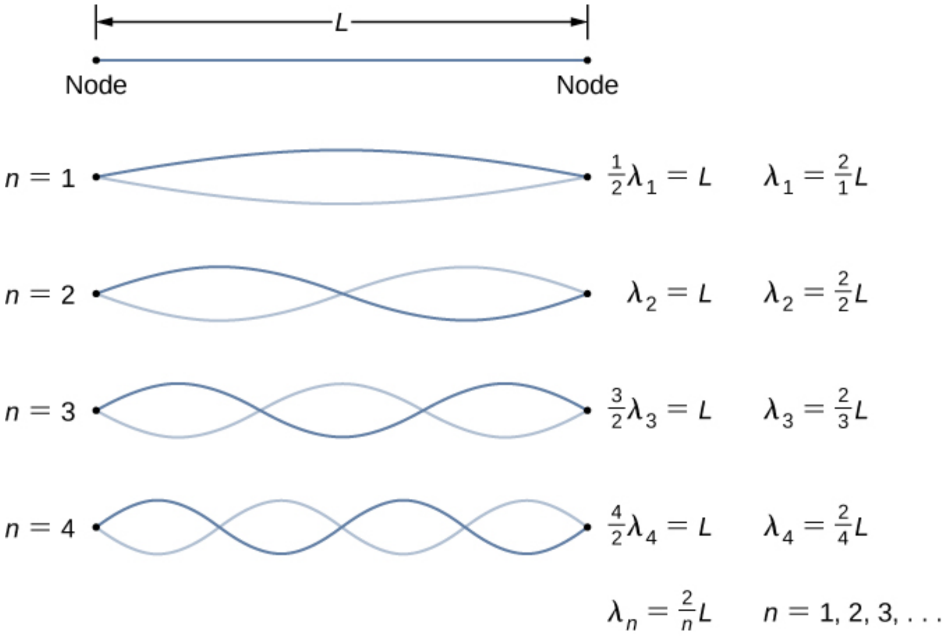
Equation:

$$f_1=\frac{v}{\lambda_1}=\frac{v}{2L},$$

where the speed of the wave is $v=\sqrt{\frac{F_T}{\mu}}$. Keeping the tension constant and increasing the frequency leads to the second harmonic or the $n=2$ mode. This mode is a full wavelength $\lambda_2=L$ and the frequency is twice the fundamental frequency:

Equation:

$$f_2=\frac{v}{\lambda_2}=\frac{v}{L}=2f_1.$$



Standing waves created on a string of length *L*. A node occurs at each end of the string. The nodes are boundary conditions that limit the possible frequencies that excite standing waves. (Note that the amplitudes of the oscillations have been kept constant for visualization. The standing wave patterns possible on the string are known as the normal modes. Conducting this experiment in the lab would result in a decrease in amplitude as the frequency increases.)

The next two modes, or the third and fourth harmonics, have wavelengths of $\lambda_3 = \frac{2}{3}L$ and $\lambda_4 = \frac{2}{4}L$, driven by frequencies of $f_3 = \frac{3v}{2L} = 3f_1$ and $f_4 = \frac{4v}{2L} = 4f_1$. All frequencies above the frequency f_1 are known as the **overtones**. The equations for the wavelength and the frequency can be summarized as:

Note:
Equation:

$$\lambda_n = \frac{2}{n}L \quad n = 1, 2, 3, 4, 5...$$

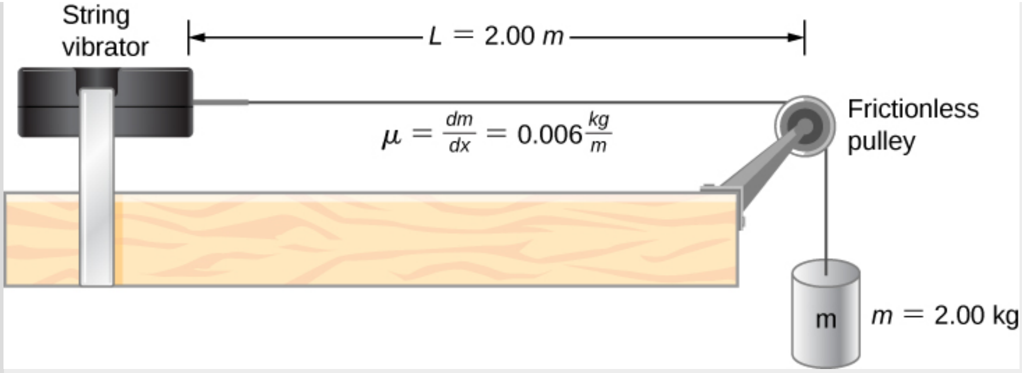
Note:
Equation:

$$f_n = n\frac{v}{2L} = nf_1 \quad n = 1, 2, 3, 4, 5...$$

The standing wave patterns that are possible for a string, the first four of which are shown in [link], are known as the **normal modes**, with frequencies known as the normal frequencies. In summary, the first frequency to produce a normal mode is called the fundamental frequency (or first harmonic). Any frequencies above the fundamental frequency are overtones. The second frequency of the $n = 2$ normal mode of the string is the first overtone (or second harmonic). The frequency of the $n = 3$ normal mode is the second overtone (or third harmonic) and so on.

The solutions shown as [link] and [link] are for a string with the boundary condition of a node on each end. When the boundary condition on either side is the same, the system is said to have symmetric boundary conditions. [link] and [link] are good for any symmetric boundary conditions, that is, nodes at both ends or antinodes at both ends.

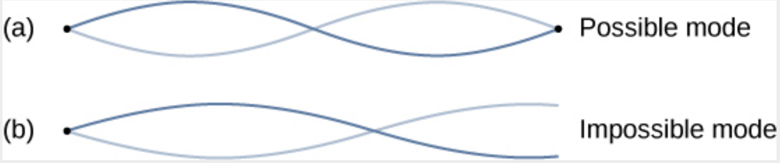
Example:
Standing Waves on a String
Consider a string of $L = 2.00\text{ m}$. attached to an adjustable-frequency string vibrator as shown in [link]. The waves produced by the vibrator travel down the string and are reflected by the fixed boundary condition at the pulley. The string, which has a linear mass density of $\mu = 0.006\text{ kg/m}$, is passed over a frictionless pulley of a negligible mass, and the tension is provided by a 2.00-kg hanging mass. (a) What is the velocity of the waves on the string? (b) Draw a sketch of the first three normal modes of the standing waves that can be produced on the string and label each with the wavelength. (c) List the frequencies that the string vibrator must be tuned to in order to produce the first three normal modes of the standing waves.



A string attached to an adjustable-frequency string vibrator.

Strategy

a. The velocity of the wave can be found using $v = \sqrt{\frac{F_T}{\mu}}$. The tension is provided by the weight of the hanging mass.
b. The standing waves will depend on the boundary conditions. There must be a node at each end. The first mode will be one half of a wave. The second can be found by adding a half wavelength. That is the shortest length that will result in a node at the boundaries. For example, adding one quarter of a wavelength will result in an antinode at the boundary and is not a mode which would satisfy the boundary conditions. This is shown in [link].
c. Since the wave speed velocity is the wavelength times the frequency, the frequency is wave speed divided by the wavelength.



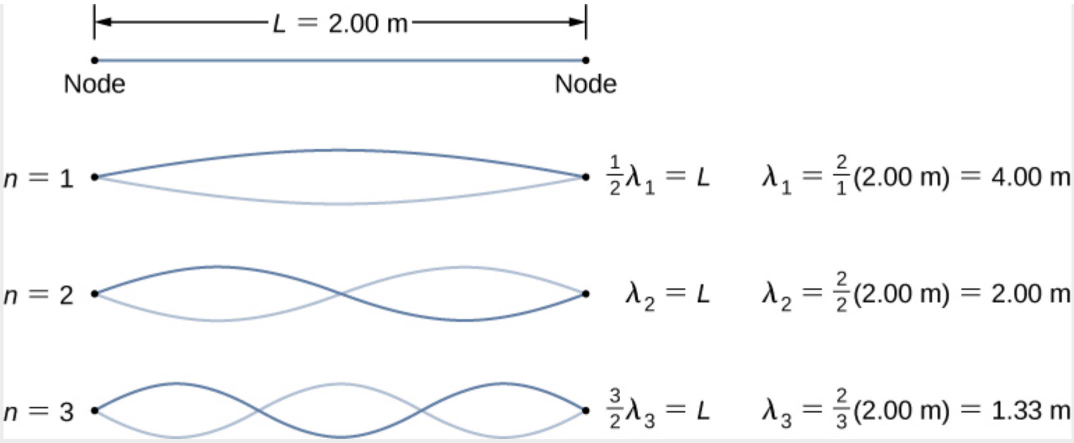
(a) The figure represents the second mode of the string that satisfies the boundary conditions of a node at each end of the string. (b)This figure could not possibly be a normal mode on the string because it does not satisfy the boundary conditions. There is a node on one end, but an antinode on the other.

Solution

a. Begin with the velocity of a wave on a string. The tension is equal to the weight of the hanging mass. The linear mass density and mass of the hanging mass are given:
Equation:

$$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{mg}{\mu}} = \sqrt{\frac{2\text{ kg}\left(9.8\frac{\text{m}}{\text{s}}\right)}{0.006\frac{\text{kg}}{\text{m}}}} = 57.15\text{ m/s}.$$

b. The first normal mode that has a node on each end is a half wavelength. The next two modes are found by adding a half of a wavelength.


$L = 2.00$ m
Node
Node
$n = 1$
$\frac{1}{2}\lambda_1 = L$
$\lambda_1 = \frac{2}{1}(2.00\text{ m}) = 4.00\text{ m}$
$n = 2$
$\lambda_2 = L$
$\lambda_2 = \frac{2}{2}(2.00\text{ m}) = 2.00\text{ m}$
$n = 3$
$\frac{3}{2}\lambda_3 = L$
$\lambda_3 = \frac{2}{3}(2.00\text{ m}) = 1.33\text{ m}$

c. The frequencies of the first three modes are found by using $f = \frac{v_w}{\lambda}$.

Equation:

$$f_1 = \frac{v_w}{\lambda_1} = \frac{57.15\text{ m/s}}{4.00\text{ m}} = 14.29\text{ Hz}$$
$$f_2 = \frac{v_w}{\lambda_2} = \frac{57.15\text{ m/s}}{2.00\text{ m}} = 28.58\text{ Hz}$$
$$f_3 = \frac{v_w}{\lambda_3} = \frac{57.15\text{ m/s}}{1.333\text{ m}} = 42.87\text{ Hz}$$

Significance

The three standing modes in this example were produced by maintaining the tension in the string and adjusting the driving frequency. Keeping the tension in the string constant results in a constant velocity. The same modes could have been produced by keeping the frequency constant and adjusting the speed of the wave in the string (by changing the hanging mass.)

Note:

Visit this simulation to play with a 1D or 2D system of coupled mass-spring oscillators. Vary the number of masses, set the initial conditions, and watch the system evolve. See the spectrum of normal modes for arbitrary motion. See longitudinal or transverse modes in the 1D system.

Note:

Exercise:

Problem:

Check Your Understanding The equations for the wavelengths and the frequencies of the modes of a wave produced on a string:

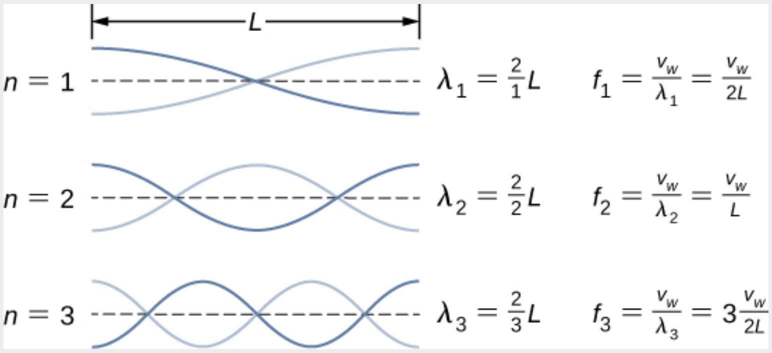
Equation:

$$\lambda_n = \frac{2}{n}L \quad n = 1, 2, 3, 4, 5\text{... and}$$
$$f_n = n\frac{v}{2L} = nf_1 \quad n = 1, 2, 3, 4, 5\text{...}$$

were derived by considering a wave on a string where there were symmetric boundary conditions of a node at each end. These modes resulted from two sinusoidal waves with identical characteristics except they were moving in opposite directions, confined to a region *L* with nodes required at both ends. Will the same equations work if there were symmetric boundary conditions with antinodes at each end? What would the normal modes look like for a medium that was free to oscillate on each end? Don't worry for now if you cannot imagine such a medium, just consider two sinusoidal wave functions in a region of length *L*, with antinodes on each end.

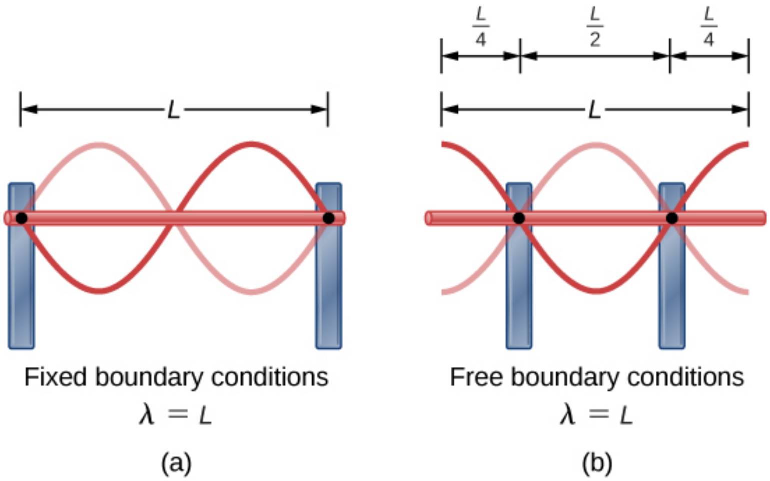
Solution:

Yes, the equations would work equally well for symmetric boundary conditions of a medium free to oscillate on each end where there was an antinode on each end. The normal modes of the first three modes are shown below. The dotted line shows the equilibrium position of the medium.

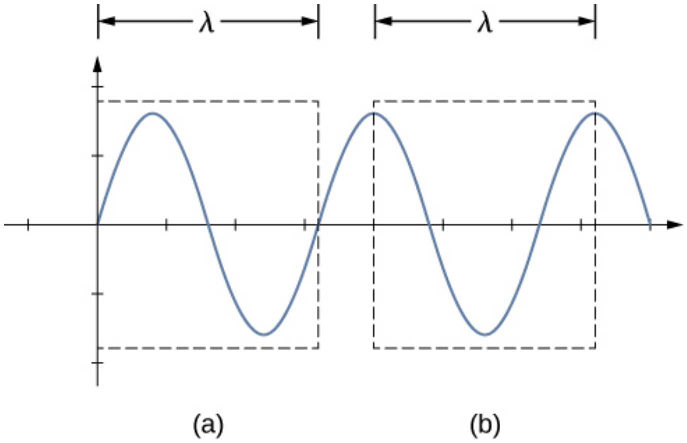

L
$n = 1$
$\lambda_1 = \frac{2}{1}L$
$f_1 = \frac{v_w}{\lambda_1} = \frac{v_w}{2L}$
$n = 2$
$\lambda_2 = \frac{2}{2}L$
$f_2 = \frac{v_w}{\lambda_2} = \frac{v_w}{L}$
$n = 3$
$\lambda_3 = \frac{2}{3}L$
$f_3 = \frac{v_w}{\lambda_3} = 3\frac{v_w}{2L}$

Note that the first mode is two quarters, or one half, of a wavelength. The second mode is one quarter of a wavelength, followed by one half of a wavelength, followed by one quarter of a wavelength, or one full wavelength. The third mode is one and a half wavelengths. These are the same result as the string with a node on each end. The equations for symmetrical boundary conditions work equally well for fixed boundary conditions and free boundary conditions. These results will be revisited in the next chapter when discussing sound wave in an open tube.

The free boundary conditions shown in the last Check Your Understanding may seem hard to visualize. How can there be a system that is free to oscillate on each end? In [link] are shown two possible configuration of a metallic rods (shown in red) attached to two supports (shown in blue). In part (a), the rod is supported at the ends, and there are fixed boundary conditions at both ends. Given the proper frequency, the rod can be driven into resonance with a wavelength equal to length of the rod, with nodes at each end. In part (b), the rod is supported at positions one quarter of the length from each end of the rod, and there are free boundary conditions at both ends. Given the proper frequency, this rod can also be driven into resonance with a wavelength equal to the length of the rod, but there are antinodes at each end. If you are having trouble visualizing the wavelength in this figure, remember that the wavelength may be measured between any two nearest identical points and consider [link].



(a) A metallic rod of length *L* (red) supported by two supports (blue) on each end. When driven at the proper frequency, the rod can resonate with a wavelength equal to the length of the rod with a node on each end. (b) The same metallic rod of length *L* (red) supported by two supports (blue) at a position a quarter of the length of the rod from each end. When driven at the proper frequency, the rod can resonate with a wavelength equal to the length of the rod with an antinode on each end.



A wavelength may be measure between the nearest two repeating points. On the wave on a string, this means the same height and slope. (a) The wavelength is measured between the two nearest points where the height is zero and the slope is maximum and positive. (b) The wavelength is measured between two identical points where the height is maximum and the slope is zero.

Note that the study of standing waves can become quite complex. In [link](a), the $n=2$ mode of the standing wave is shown, and it results in a wavelength equal to L. In this configuration, the $n=1$ mode would also have been possible with a standing wave equal to $2L$. Is it possible to get the $n=1$ mode for the configuration shown in part (b)? The answer is no. In this configuration, there are additional conditions set beyond the boundary conditions. Since the rod is mounted at a point one quarter of the length from each side, a node must exist there, and this limits the possible modes of standing waves that can be created. We leave it as an exercise for the reader to consider if other modes of standing waves are possible. It should be noted that when a system is driven at a frequency that does not cause the system to resonate, vibrations may still occur, but the amplitude of the vibrations will be much smaller than the amplitude at resonance.

A field of mechanical engineering uses the sound produced by the vibrating parts of complex mechanical systems to troubleshoot problems with the systems. Suppose a part in an automobile is resonating at the frequency of the car’s engine, causing unwanted vibrations in the automobile. This may cause the engine to fail prematurely. The engineers use microphones to record the sound produced by the engine, then use a technique called Fourier analysis to find frequencies of sound produced with large amplitudes and then look at the parts list of the automobile to find a part that would resonate at that frequency. The solution may be as simple as changing the composition of the material used or changing the length of the part in question.

There are other numerous examples of resonance in standing waves in the physical world. The air in a tube, such as found in a musical instrument like a flute, can be forced into resonance and produce a pleasant sound, as we discuss in Sound.

At other times, resonance can cause serious problems. A closer look at earthquakes provides evidence for conditions appropriate for resonance, standing waves, and constructive and destructive interference. A building may vibrate for several seconds with a driving frequency matching that of the natural frequency of vibration of the building—producing a resonance resulting in one building collapsing while neighboring buildings do not. Often, buildings of a certain height are devastated while other taller buildings remain intact. The building height matches the condition for setting up a standing wave for that particular height. The span of the roof is also important. Often it is seen that gymnasiums, supermarkets, and churches suffer damage when individual homes suffer far less damage. The roofs with large surface areas supported only at the edges resonate at the frequencies of the earthquakes, causing them to collapse. As the earthquake waves travel along the surface of Earth and reflect off denser rocks, constructive interference occurs at certain points. Often areas closer to the epicenter are not damaged, while areas farther away are damaged.

Summary

- A standing wave is the superposition of two waves which produces a wave that varies in amplitude but does not propagate.
- Nodes are points of no motion in standing waves.
- An antinode is the location of maximum amplitude of a standing wave.
- Normal modes of a wave on a string are the possible standing wave patterns. The lowest frequency that will produce a standing wave is known as the fundamental frequency. The higher frequencies which produce standing waves are called overtones.

Key Equations

Wave speed	$v = \frac{\lambda}{T} = \lambda f$

Linear mass density	$\mu = \frac{\text{mass of the string}}{\text{length of the string}}$
Speed of a wave or pulse on a string under tension	$\lvert v \rvert = \sqrt{\frac{F_T}{\mu}}$
Speed of a compression wave in a fluid	$v = \sqrt{\frac{B}{\rho}}$
Resultant wave from superposition of two sinusoidal waves that are identical except for a phase shift	$y_R\left(x, t\right) = \left[2A \cos\left(\frac{\phi}{2}\right)\right] \sin\left(kx - \omega t + \frac{\phi}{2}\right)$
Wave number	$k \equiv \frac{2\pi}{\lambda}$
Wave speed	$v = \frac{\omega}{k}$
A periodic wave	$y\left(x, t\right) = A \sin\left(kx \mp \omega t + \phi\right)$
Phase of a wave	$kx \mp \omega t + \phi$
The linear wave equation	$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v_w^2} \frac{\partial^2 y(x,t)}{\partial t^2}$
Power averaged over a wavelength	$P_{\text{ave}} = \frac{E_\lambda}{T} = \frac{1}{2}\mu A^2 \omega^2 \frac{\lambda}{T} = \frac{1}{2}\mu A^2 \omega^2 v$
Intensity	$I = \frac{P}{A}$
Intensity for a spherical wave	$I = \frac{P}{4\pi r^2}$
Equation of a standing wave	$y\left(x, t\right) = \left[2A \sin\left(kx\right)\right] \cos\left(\omega t\right)$
Wavelength for symmetric boundary conditions	$\lambda_n = \frac{2}{n} L, \quad n = 1, 2, 3, 4, 5 \ldots$
Frequency for symmetric boundary conditions	$f_n = n \frac{v}{2L} = n f_1, \quad n = 1, 2, 3, 4, 5 \ldots$

Conceptual Questions

Exercise:

Problem:

A truck manufacturer finds that a strut in the engine is failing prematurely. A sound engineer determines that the strut resonates at the frequency of the engine and suspects that this could be the problem. What are two possible characteristics of the strut can be modified to correct the problem?

Solution:

It may be as easy as changing the length and/or the density a small amount so that the parts do not resonate at the frequency of the motor.

Exercise:

Problem:

Why do roofs of gymnasiums and churches seem to fail more than family homes when an earthquake occurs?

Exercise:

Problem:

Wine glasses can be set into resonance by moistening your finger and rubbing it around the rim of the glass. Why?

Solution:

Energy is supplied to the glass by the work done by the force of your finger on the glass. When supplied at the right frequency, standing waves form. The glass resonates and the vibrations produce sound.

Exercise:

Problem:

Air conditioning units are sometimes placed on the roof of homes in the city. Occasionally, the air conditioners cause an undesirable hum throughout the upper floors of the homes. Why does this happen? What can be done to reduce the hum?

Exercise:

Problem:

Consider a standing wave modeled as $y\left(x,t\right)=4.00\;\text{cm}\;\text{sin}\left(3\;{\text{m}}^{-1}x\right)\text{cos}\left(4\;{\text{s}}^{-1}t\right)$. Is there a node or an antinode at $x=0.00\;\text{m}$? What about a standing wave modeled as $y\left(x,t\right)=4.00\;\text{cm}\;\text{sin}\left(3\;{\text{m}}^{-1}x+\frac{\pi}{2}\right)\text{cos}\left(4\;{\text{s}}^{-1}t\right)$? Is there a node or an antinode at the $x=0.00\;\text{m}$ position?

Solution:

For the equation $y\left(x,t\right)=4.00\;\text{cm}\;\text{sin}\left(3\;{\text{m}}^{-1}x\right)\text{cos}\left(4\;{\text{s}}^{-1}t\right)$, there is a node because when $x=0.00\;\text{m}$, $\text{sin}\left(3\;{\text{m}}^{-1}\left(0.00\;\text{m}\right)\right)=0.00$, so $y\left(0.00\;\text{m},t\right)=0.00\;\text{m}$ for all time. For the equation $y\left(x,t\right)=4.00\;\text{cm}\;\text{sin}\left(3\;{\text{m}}^{-1}x+\frac{\pi}{2}\right)\text{cos}\left(4\;{\text{s}}^{-1}t\right)$, there is an antinode because when $x=0.00\;\text{m}$, $\text{sin}\left(3\;{\text{m}}^{-1}\left(0.00\;\text{m}\right)+\frac{\pi}{2}\right)=+1.00$, so $y\left(0.00\;\text{m},t\right)$ oscillates between $+A$ and $-A$ as the cosine term oscillates between +1 and -1.

Problems

Exercise:

Problem:

A wave traveling on a Slinky® that is stretched to 4 m takes 2.4 s to travel the length of the Slinky and back again. (a) What is the speed of the wave? (b) Using the same Slinky stretched to the same length, a standing wave is created which consists of three antinodes and four nodes. At what frequency must the Slinky be oscillating?

Exercise:

Problem:

A 2-m long string is stretched between two supports with a tension that produces a wave speed equal to $v_w = 50.00\,\text{m/s}$. What are the wavelength and frequency of the first three modes that resonate on the string?

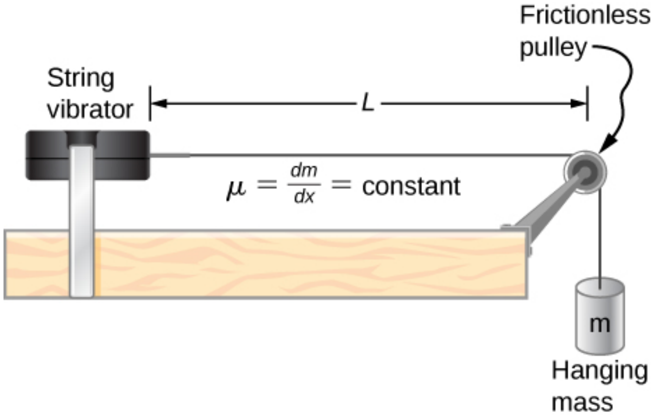
Solution:

$$\begin{aligned}
&\lambda_n = \tfrac{2.00}{n}L, \quad f_n = \tfrac{v}{\lambda_n} \\
&\lambda_1 = 4.00\,\text{m}, \quad f_1 = 12.5\,\text{Hz} \\
&\lambda_2 = 2.00\,\text{m}, \quad f_2 = 25.00\,\text{Hz} \\
&\lambda_3 = 1.33\,\text{m}, \quad f_3 = 37.59\,\text{Hz}
\end{aligned}$$

Exercise:

Problem:

Consider the experimental setup shown below. The length of the string between the string vibrator and the pulley is $L = 1.00\,\text{m}$. The linear density of the string is $\mu = 0.006\,\text{kg/m}$. The string vibrator can oscillate at any frequency. The hanging mass is 2.00 kg. (a)What are the wavelength and frequency of $n = 6$ mode? (b) The string oscillates the air around the string. What is the wavelength of the sound if the speed of the sound is $v_s = 343.00\,\text{m/s}$?



Exercise:

Problem:

A cable with a linear density of $\mu = 0.2\,\text{kg/m}$ is hung from telephone poles. The tension in the cable is 500.00 N. The distance between poles is 20 meters. The wind blows across the line, causing the cable resonate. A standing waves pattern is produced that has 4.5 wavelengths between the two poles. The speed of sound at the current temperature $T = 20^\circ\text{C}$ is $343.00\,\text{m/s}$. What are the frequency and wavelength of the hum?

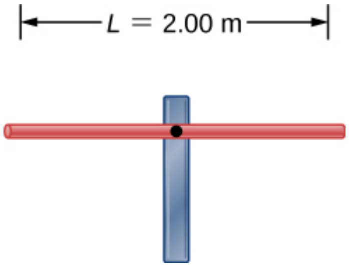
Solution:

$$\begin{aligned}
&v = 158.11\,\text{m/s}, \quad \lambda = 4.44\,\text{m}, \quad f = 35.61\,\text{Hz} \\
&\lambda_s = 9.63\,\text{m}
\end{aligned}$$

Exercise:

Problem:

Consider a rod of length L, mounted in the center to a support. A node must exist where the rod is mounted on a support, as shown below. Draw the first two normal modes of the rod as it is driven into resonance. Label the wavelength and the frequency required to drive the rod into resonance.



Exercise:

Problem:

Consider two wave functions $y\left(x,t\right)=0.30\ \text{cm}\ \sin\left(3\ \text{m}^{-1}x-4\ \text{s}^{-1}t\right)$ and $y\left(x,t\right)=0.30\ \text{cm}\ \sin\left(3\ \text{m}^{-1}x+4\ \text{s}^{-1}t\right)$. Write a wave function for the resulting standing wave.

Solution:

$y\left(x,t\right)=\left[0.60\ \text{cm}\ \sin\left(3\ \text{m}^{-1}x\right)\right]\cos\left(4\ \text{s}^{-1}t\right)$

Exercise:

Problem:

A 2.40-m wire has a mass of 7.50 g and is under a tension of 160 N. The wire is held rigidly at both ends and set into oscillation. (a) What is the speed of waves on the wire? The string is driven into resonance by a frequency that produces a standing wave with a wavelength equal to 1.20 m. (b) What is the frequency used to drive the string into resonance?

Exercise:

Problem:

A string with a linear mass density of 0.0062 kg/m and a length of 3.00 m is set into the $n=100$ mode of resonance. The tension in the string is 20.00 N. What is the wavelength and frequency of the wave?

Solution:

$\lambda_{100}=0.06\ \text{m}$

$v=56.8\ \text{m/s},\quad f_n=nf_1,\quad n=1,2,3,4,5...$

$f_{100}=947\ \text{Hz}$

Exercise:

Problem:

A string with a linear mass density of 0.0075 kg/m and a length of 6.00 m is set into the $n=4$ mode of resonance by driving with a frequency of 100.00 Hz. What is the tension in the string?

Exercise:

Problem:

Two sinusoidal waves with identical wavelengths and amplitudes travel in opposite directions along a string producing a standing wave. The linear mass density of the string is $\mu = 0.075\,\mathrm{kg/m}$ and the tension in the string is $F_T = 5.00\,\mathrm{N}$. The time interval between instances of total destructive interference is $\Delta t = 0.13\,\mathrm{s}$. What is the wavelength of the waves?

Solution:

$T = 2\Delta t, \quad v = \frac{\lambda}{T}, \quad \lambda = 2.12\,\mathrm{m}$

Exercise:

Problem:

A string, fixed on both ends, is 5.00 m long and has a mass of 0.15 kg. The tension if the string is 90 N. The string is vibrating to produce a standing wave at the fundamental frequency of the string. (a) What is the speed of the waves on the string? (b) What is the wavelength of the standing wave produced? (c) What is the period of the standing wave?

Exercise:

Problem:

A string is fixed at both end. The mass of the string is 0.0090 kg and the length is 3.00 m. The string is under a tension of 200.00 N. The string is driven by a variable frequency source to produce standing waves on the string. Find the wavelengths and frequency of the first four modes of standing waves.

Solution:

$$\begin{aligned}
&\lambda_1 = 6.00\,\mathrm{m}, \quad \lambda_2 = 3.00\,\mathrm{m}, \quad \lambda_3 = 2.00\,\mathrm{m}, \quad \lambda_4 = 1.50\,\mathrm{m} \\
&v = 258.20\,\mathrm{m/s} = \lambda f \\
&f_1 = 43.03\,\mathrm{Hz}, \quad f_2 = 86.07\,\mathrm{Hz}, \quad f_3 = 129.10\,\mathrm{Hz}, \quad f_4 = 172.13\,\mathrm{Hz}
\end{aligned}$$

Exercise:

Problem:

The frequencies of two successive modes of standing waves on a string are 258.36 Hz and 301.42 Hz. What is the next frequency above 100.00 Hz that would produce a standing wave?

Exercise:

Problem:

A string is fixed at both ends to supports 3.50 m apart and has a linear mass density of $\mu = 0.005\,\mathrm{kg/m}$. The string is under a tension of 90.00 N. A standing wave is produced on the string with six nodes and five antinodes. What are the wave speed, wavelength, frequency, and period of the standing wave?

Solution:

$v = 134.16\,\mathrm{ms}, \lambda = 1.4\,\mathrm{m}, f = 95.83\,\mathrm{Hz}, T = 0.0104\,\mathrm{s}$

Exercise:

Problem:

Sine waves are sent down a 1.5-m-long string fixed at both ends. The waves reflect back in the opposite direction. The amplitude of the wave is 4.00 cm. The propagation velocity of the waves is 175 m/s. The $n = 6$ resonance mode of the string is produced. Write an equation for the resulting standing wave.

Additional Problems

Exercise:

Problem:

Ultrasound equipment used in the medical profession uses sound waves of a frequency above the range of human hearing. If the frequency of the sound produced by the ultrasound machine is $f = 30\,\text{kHz}$, what is the wavelength of the ultrasound in bone, if the speed of sound in bone is $v = 3000\,\text{m/s}$?

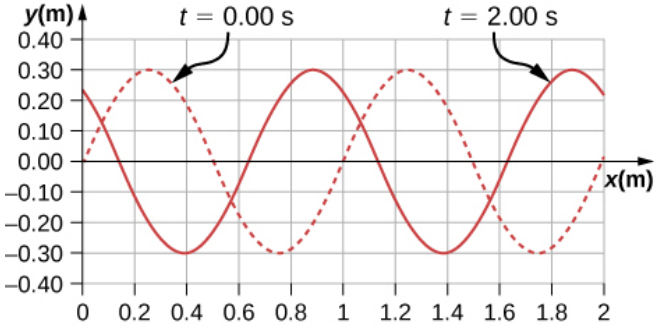
Solution:

$\lambda = 0.10\,\text{m}$

Exercise:

Problem:

Shown below is the plot of a wave function that models a wave at time $t = 0.00\,\text{s}$ and $t = 2.00\,\text{s}$. The dotted line is the wave function at time $t = 0.00\,\text{s}$ and the solid line is the function at time $t = 2.00\,\text{s}$. Estimate the amplitude, wavelength, velocity, and period of the wave.



Exercise:

Problem:

The speed of light in air is approximately $v = 3.00 \times 10^8\,\text{m/s}$ and the speed of light in glass is $v = 2.00 \times 10^8\,\text{m/s}$. A red laser with a wavelength of $\lambda = 633.00\,\text{nm}$ shines light incident of the glass, and some of the red light is transmitted to the glass. The frequency of the light is the same for the air and the glass. (a) What is the frequency of the light? (b) What is the wavelength of the light in the glass?

Solution:

a. $f = 4.74 \times 10^{14}\,\text{Hz}$; b. $\lambda = 422\,\text{nm}$

Exercise:

Problem:

A radio station broadcasts radio waves at a frequency of 101.7 MHz. The radio waves move through the air at approximately the speed of light in a vacuum. What is the wavelength of the radio waves?

Exercise:

Problem:

A sunbather stands waist deep in the ocean and observes that six crests of periodic surface waves pass each minute. The crests are 16.00 meters apart. What is the wavelength, frequency, period, and speed of the waves?

Solution:

$\lambda = 16.00\,\text{m}, \quad f = 0.10\,\text{Hz}, \quad T = 10.00\,\text{s}, \quad v = 1.6\,\text{m/s}$

Exercise:

Problem:

A tuning fork vibrates producing sound at a frequency of 512 Hz. The speed of sound of sound in air is $v = 343.00\,\text{m/s}$ if the air is at a temperature of $20.00\,^\circ\text{C}$. What is the wavelength of the sound?

Exercise:

Problem:

A motorboat is traveling across a lake at a speed of $v_b = 15.00\,\text{m/s}$. The boat bounces up and down every 0.50 s as it travels in the same direction as a wave. It bounces up and down every 0.30 s as it travels in a direction opposite the direction of the waves. What is the speed and wavelength of the wave?

Solution:

$\lambda = (v_b + v)t_b, \quad v = 3.75\,\text{m/s}, \quad \lambda = 3.00\,\text{m}$

Exercise:

Problem:

Use the linear wave equation to show that the wave speed of a wave modeled with the wave function $y\left(x, t\right) = 0.20\,\text{m}\,\sin\left(3.00\,\text{m}^{-1}x + 6.00\,\text{s}^{-1}t\right)$ is $v = 2.00\,\text{m/s}$. What are the wavelength and the speed of the wave?

Exercise:

Problem:

Given the wave functions $y_1\left(x, t\right) = A\sin\left(kx - \omega t\right)$ and $y_2\left(x, t\right) = A\sin\left(kx - \omega t + \phi\right)$ with $\phi \neq \frac{\pi}{2}$, show that $y_1\left(x, t\right) + y_2\left(x, t\right)$ is a solution to the linear wave equation with a wave velocity of $v = \frac{\omega}{k}$.

Solution:

$$\frac{\partial^2(y_1+y_2)}{\partial t^2} = -A\omega^2 \sin(kx-\omega t) - A\omega^2 \sin(kx-\omega t+\phi)$$
$$\frac{\partial^2(y_1+y_2)}{\partial x^2} = -Ak^2 \sin(kx-\omega t) - Ak^2 \sin(kx-\omega t+\phi)$$
$$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2}$$
$$-A\omega^2 \sin(kx-\omega t) - A\omega^2 \sin(kx-\omega t+\phi) = \left(\frac{1}{v^2}\right)\left(-Ak^2 \sin(kx-\omega t) - Ak^2 \sin(kx-\omega t+\phi)\right)$$
$$v = \frac{\omega}{k}$$

Exercise:

Problem:

A transverse wave on a string is modeled with the wave function $y(x,t) = 0.10\,\mathrm{m}\sin\left(0.15\,\mathrm{m}^{-1}x + 1.50\,\mathrm{s}^{-1}t + 0.20\right)$. (a) Find the wave velocity. (b) Find the position in the *y*-direction, the velocity perpendicular to the motion of the wave, and the acceleration perpendicular to the motion of the wave, of a small segment of the string centered at $x = 0.40\,\mathrm{m}$ at time $t = 5.00\,\mathrm{s}$.

Exercise:

Problem:

A sinusoidal wave travels down a taut, horizontal string with a linear mass density of $\mu = 0.060\,\mathrm{kg/m}$. The magnitude of maximum vertical acceleration of the wave is $a_{y\,\mathrm{max}} = 0.90\,\mathrm{cm/s}^2$ and the amplitude of the wave is 0.40 m. The string is under a tension of $F_T = 600.00\,\mathrm{N}$. The wave moves in the negative *x*-direction. Write an equation to model the wave.

Solution:

$$y(x,t) = 0.40\,\mathrm{m}\sin\left(0.015\,\mathrm{m}^{-1}x + 1.5\,\mathrm{s}^{-1}t\right)$$

Exercise:

Problem:

A transverse wave on a string $(\mu = 0.0030\,\mathrm{kg/m})$ is described with the equation $y(x,t) = 0.30\,\mathrm{m}\sin\left(\frac{2\pi}{4.00\,\mathrm{m}}\left(x - 16.00\frac{\mathrm{m}}{\mathrm{s}}t\right)\right)$. What is the tension under which the string is held taut?

Exercise:

Problem:

A transverse wave on a horizontal string $(\mu = 0.0060\,\mathrm{kg/m})$ is described with the equation $y(x,t) = 0.30\,\mathrm{m}\sin\left(\frac{2\pi}{4.00\,\mathrm{m}}(x - v_w t)\right)$. The string is under a tension of 300.00 N. What are the wave speed, wave number, and angular frequency of the wave?

Solution:

$$v = 223.61\,\mathrm{m/s},\ k = 1.57\,\mathrm{m}^{-1},\ \omega = 142.43\,\mathrm{s}^{-1}$$

Exercise:

Problem:

A student holds an inexpensive sonic range finder and uses the range finder to find the distance to the wall. The sonic range finder emits a sound wave. The sound wave reflects off the wall and returns to the range finder. The round trip takes 0.012 s. The range finder was calibrated for use at room temperature $T = 20^\circ\text{C}$, but the temperature in the room is actually $T = 23^\circ\text{C}$. Assuming that the timing mechanism is perfect, what percentage of error can the student expect due to the calibration?

Exercise:

Problem:

A wave on a string is driven by a string vibrator, which oscillates at a frequency of 100.00 Hz and an amplitude of 1.00 cm. The string vibrator operates at a voltage of 12.00 V and a current of 0.20 A. The power consumed by the string vibrator is $P = IV$. Assume that the string vibrator is 90% efficient at converting electrical energy into the energy associated with the vibrations of the string. The string is 3.00 m long, and is under a tension of 60.00 N. What is the linear mass density of the string?

Solution:

$$P = \frac{1}{2}A^2(2\pi f)^2\sqrt{\mu F_T}$$
$$\mu = 2.00\ \times\ 10^{-4}\,\text{kg/m}$$

Exercise:

Problem:

A traveling wave on a string is modeled by the wave equation $y\left(x,t\right) = 3.00\,\text{cm}\,\sin\left(8.00\,\text{m}^{-1}x + 100.00\,\text{s}^{-1}t\right)$. The string is under a tension of 50.00 N and has a linear mass density of $\mu = 0.008\,\text{kg/m}$. What is the average power transferred by the wave on the string?

Exercise:

Problem:

A transverse wave on a string has a wavelength of 5.0 m, a period of 0.02 s, and an amplitude of 1.5 cm. The average power transferred by the wave is 5.00 W. What is the tension in the string?

Solution:

$$P = \frac{1}{2}\mu A^2\omega^2\frac{\lambda}{T},\ \mu = 0.0018\,\text{kg/m}$$

Exercise:

Problem:

(a) What is the intensity of a laser beam used to burn away cancerous tissue that, when 90.0% absorbed, puts 500.0 J of energy into a circular spot 2.00 mm in diameter in 4.00 s? (b) Discuss how this intensity compares to the average intensity of sunlight (about 1kW/m^2) and the implications that would have if the laser beam entered your eye. Note how your answer depends on the time duration of the exposure.

Exercise:

Problem:

Consider two periodic wave functions, $y_1\left(x,t\right)=A\sin\left(kx-\omega t\right)$ and $y_2\left(x,t\right)=A\sin\left(kx-\omega t+\phi\right)$. (a) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of 2*A*? (b) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of zero?

Solution:

a. $A_R=2A\cos\left(\frac{\phi}{2}\right),\ \cos\left(\frac{\phi}{2}\right)=1,\ \phi=0,2\pi,4\pi,$...; b. $A_R=2A\cos\left(\frac{\phi}{2}\right),\ \cos\left(\frac{\phi}{2}\right)=0,\ \phi=0,\pi,3\pi,5\pi$...

Exercise:

Problem:

Consider two periodic wave functions, $y_1\left(x,t\right)=A\sin\left(kx-\omega t\right)$ and $y_2\left(x,t\right)=A\cos\left(kx-\omega t+\phi\right)$. (a) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of 2*A*? (b) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of zero?

Exercise:

Problem:

A trough with dimensions 10.00 meters by 0.10 meters by 0.10 meters is partially filled with water. Small-amplitude surface water waves are produced from both ends of the trough by paddles oscillating in simple harmonic motion. The height of the water waves are modeled with two sinusoidal wave equations, $y_1\left(x,t\right)=0.3\,\text{m}\sin\left(4\,\text{m}^{-1}x-3\,\text{s}^{-1}t\right)$ and $y_2\left(x,t\right)=0.3\,\text{m}\cos\left(4\,\text{m}^{-1}x+3\,\text{s}^{-1}t-\frac{\pi}{2}\right)$. What is the wave function of the resulting wave after the waves reach one another and before they reach the end of the trough (i.e., assume that there are only two waves in the trough and ignore reflections)? Use a spreadsheet to check your results. (*Hint:* Use the trig identities $\sin\left(u\pm v\right)=\sin u\cos v\pm\cos u\sin v$ and $\cos\left(u\pm v\right)=\cos u\cos v\mp\sin u\sin v$)

Solution:

$y_R\left(x,t\right)=0.6\,\text{m}\sin\left(4\,\text{m}^{-1}x\right)\cos\left(3\,\text{s}^{-1}t\right)$

Exercise:

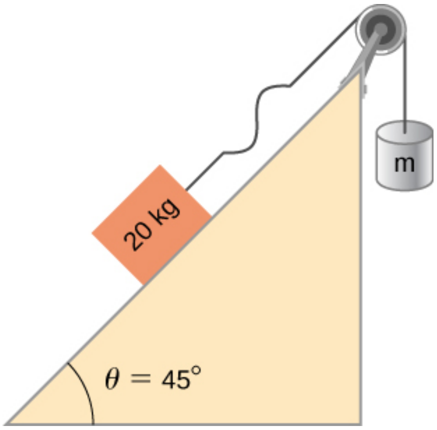
Problem:

A seismograph records the S- and P-waves from an earthquake 20.00 s apart. If they traveled the same path at constant wave speeds of $v_S=4.00\,\text{km/s}$ and $v_P=7.50\,\text{km/s}$, how far away is the epicenter of the earthquake?

Exercise:

Problem:

Consider what is shown below. A 20.00-kg mass rests on a frictionless ramp inclined at 45°. A string with a linear mass density of $\mu=0.025\,\text{kg/m}$ is attached to the 20.00-kg mass. The string passes over a frictionless pulley of negligible mass and is attached to a hanging mass (*m*). The system is in static equilibrium. A wave is induced on the string and travels up the ramp. (a) What is the mass of the hanging mass (*m*)? (b) At what wave speed does the wave travel up the string?



Solution:

a. $$\begin{array}{l}(1) F_T - 20.00\,\text{kg}\left(9.80\,\text{m/s}^2\right)\cos 45^\circ = 0 \\ (2) m\left(9.80\,\text{m/s}^2\right) - F_T = 0 \\ m = 14.14\,\text{kg}\end{array}$$; b. $$\begin{array}{rcl}F_T & = & 138.57\,\text{N} \\ v & = & 74.45\,\text{m/s}\end{array}$$

Exercise:

Problem:

Consider the superposition of three wave functions $y\left(x,t\right) = 3.00\,\text{cm}\sin\left(2\,\text{m}^{-1}x - 3\,\text{s}^{-1}t\right)$, $y\left(x,t\right) = 3.00\,\text{cm}\sin\left(6\,\text{m}^{-1}x + 3\,\text{s}^{-1}t\right)$, and $y\left(x,t\right) = 3.00\,\text{cm}\sin\left(2\,\text{m}^{-1}x - 4\,\text{s}^{-1}t\right)$. What is the height of the resulting wave at position $x = 3.00\,\text{m}$ at time $t = 10.0\,\text{s}$?

Exercise:

Problem:

A string has a mass of 150 g and a length of 3.4 m. One end of the string is fixed to a lab stand and the other is attached to a spring with a spring constant of $k_s = 100\,\text{N/m}$. The free end of the spring is attached to another lab pole. The tension in the string is maintained by the spring. The lab poles are separated by a distance that stretches the spring 2.00 cm. The string is plucked and a pulse travels along the string. What is the propagation speed of the pulse?

Solution:

$F_T = 2\,\text{N}, v = 6.73\,\text{m/s}$

Exercise:

Problem:

A standing wave is produced on a string under a tension of 70.0 N by two sinusoidal transverse waves that are identical, but moving in opposite directions. The string is fixed at $x = 0.00\,\text{m}$ and $x = 10.00\,\text{m}$. Nodes appear at $x = 0.00\,\text{m}$, 2.00 m, 4.00 m, 6.00 m, 8.00 m, and 10.00 m. The amplitude of the standing wave is 3.00 cm. It takes 0.10 s for the antinodes to make one complete oscillation. (a) What are the wave functions of the two sine waves that produce the standing wave? (b) What are the maximum velocity and acceleration of the string, perpendicular to the direction of motion of the transverse waves, at the antinodes?

Exercise:

Problem:

A string with a length of 4 m is held under a constant tension. The string has a linear mass density of $\mu = 0.006\ \text{kg/m}$. Two resonant frequencies of the string are 400 Hz and 480 Hz. There are no resonant frequencies between the two frequencies. (a) What are the wavelengths of the two resonant modes? (b) What is the tension in the string?

Solution:

a. $f_n = \frac{nv}{2L},\ v = \frac{2Lf_{n+1}}{n+1},\ \frac{n+1}{n} = \frac{2Lf_{n+1}}{2Lf_n},\ 1 + \frac{1}{n} = 1.2,\ n = 5$; b. $F_T = 245.76\ \text{N}$
$\lambda_n = \frac{2}{n}L,\ \lambda_5 = 1.6\ \text{m},\ \lambda_6 = 1.33\ \text{m}$

Challenge Problems

Exercise:

Problem:

A copper wire has a radius of 200 µm and a length of 5.0 m. The wire is placed under a tension of 3000 N and the wire stretches by a small amount. The wire is plucked and a pulse travels down the wire. What is the propagation speed of the pulse? (Assume the temperature does not change: $\left(\rho = 8.96\frac{\text{g}}{\text{cm}^3}, Y = 1.1\ \times\ 10^{11}\frac{\text{N}}{\text{m}}\right).$)

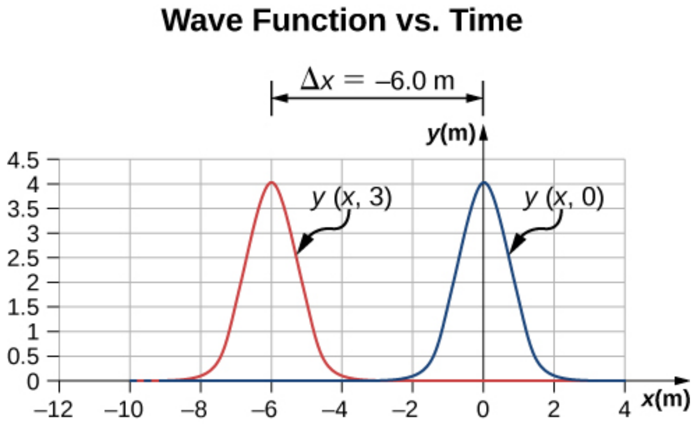
Exercise:

Problem:

A pulse moving along the *x* axis can be modeled as the wave function $y\left(x, t\right) = 4.00\ \text{m}e^{-\left(\frac{x+(2.00\ \text{m/s})t}{1.00\ \text{m}}\right)^2}$. (a)What are the direction and propagation speed of the pulse? (b) How far has the wave moved in 3.00 s? (c) Plot the pulse using a spreadsheet at time $t = 0.00\ \text{s}$ and $t = 3.00\ \text{s}$ to verify your answer in part (b).

Solution:

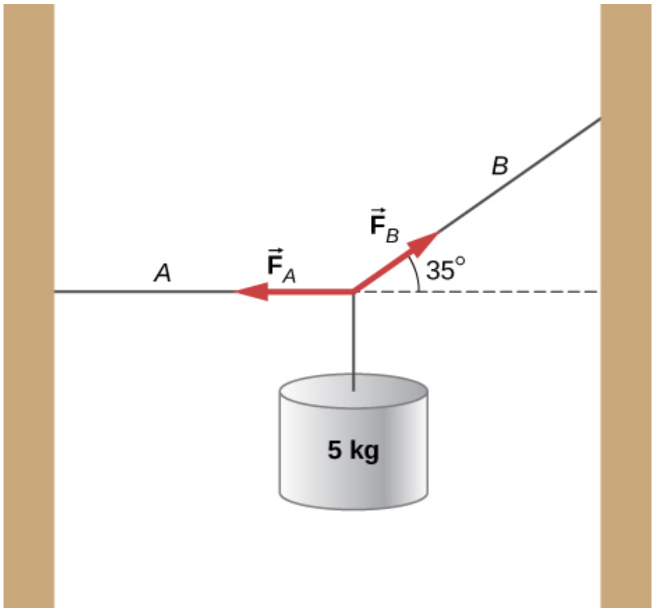
a. Moves in the negative *x* direction at a propagation speed of $v = 2.00\ \text{m/s}$. b. $\Delta x = -6.00\ \text{m}$; c.



Exercise:

Problem:

A string with a linear mass density of $\mu = 0.0085\,\text{kg/m}$ is fixed at both ends. A 5.0-kg mass is hung from the string, as shown below. If a pulse is sent along section A, what is the wave speed in section A and the wave speed in section B?



Exercise:

Problem:

Consider two wave functions $y_1(x,t) = A\sin(kx - \omega t)$ and $y_2(x,t) = A\sin(kx + \omega t + \phi)$. What is the wave function resulting from the interference of the two wave? (*Hint:* $\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$ and $\phi = \frac{\phi}{2} + \frac{\phi}{2}$.)

Solution:

$$\sin(kx - \omega t) = \sin\left(kx + \frac{\phi}{2}\right)\cos\left(\omega t + \frac{\phi}{2}\right) - \cos\left(kx + \frac{\phi}{2}\right)\sin\left(\omega t + \frac{\phi}{2}\right)$$

$$\sin(kx - \omega t + \phi) = \sin\left(kx + \frac{\phi}{2}\right)\cos\left(\omega t + \frac{\phi}{2}\right) + \cos\left(kx + \frac{\phi}{2}\right)\sin\left(\omega t + \frac{\phi}{2}\right)$$

$$\sin(kx - \omega t) + \sin(kx + \omega t + \phi) = 2\sin\left(kx + \frac{\phi}{2}\right)\cos\left(\omega t + \frac{\phi}{2}\right)$$

$$y_R = 2\,A\sin\left(kx + \frac{\phi}{2}\right)\cos\left(\omega t + \frac{\phi}{2}\right)$$

Exercise:

Problem:

The wave function that models a standing wave is given as $y_R(x,t) = 6.00\,\text{cm}\sin\left(3.00\,\text{m}^{-1}x + 1.20\,\text{rad}\right)\cos\left(6.00\,\text{s}^{-1}t + 1.20\,\text{rad}\right)$. What are two wave functions that interfere to form this wave function? Plot the two wave functions and the sum of the sum of the two wave functions at $t = 1.00\,\text{s}$ to verify your answer.

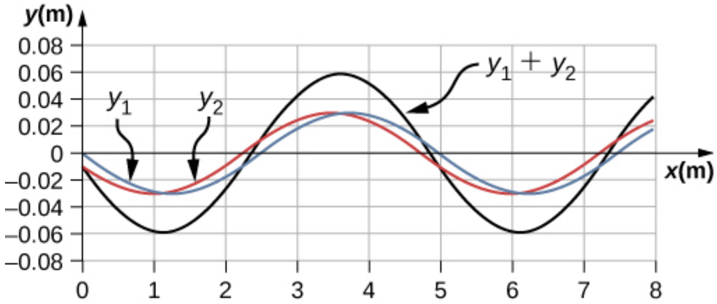
Exercise:

Problem:

Consider two wave functions $y_1\left(x,t\right) = A\sin\left(kx - \omega t\right)$ and $y_2\left(x,t\right) = A\sin\left(kx + \omega t + \phi\right)$. The resultant wave form when you add the two functions is $y_R = 2A\sin\left(kx + \frac{\phi}{2}\right)\cos\left(\omega t + \frac{\phi}{2}\right)$. Consider the case where $A = 0.03\,\text{m}^{-1}$, $k = 1.26\,\text{m}^{-1}$, $\omega = \pi\,\text{s}^{-1}$, and $\phi = \frac{\pi}{10}$. (a) Where are the first three nodes of the standing wave function starting at zero and moving in the positive *x* direction? (b) Using a spreadsheet, plot the two wave functions and the resulting function at time $t = 1.00\,\text{s}$ to verify your answer.

Solution:

$\sin\left(kx + \frac{\phi}{2}\right) = 0,\ kx + \frac{\phi}{2} = 0, \pi, 2\pi,\ 1.26\,\text{m}^{-1}x + \frac{\pi}{20} = \pi, 2\pi, 3\pi;$
$x = 2.37\,\text{m}, 4.86\,\text{m}, 7.35\,\text{m}$



Glossary

antinode
    location of maximum amplitude in standing waves

fundamental frequency
    lowest frequency that will produce a standing wave

node
    point where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave

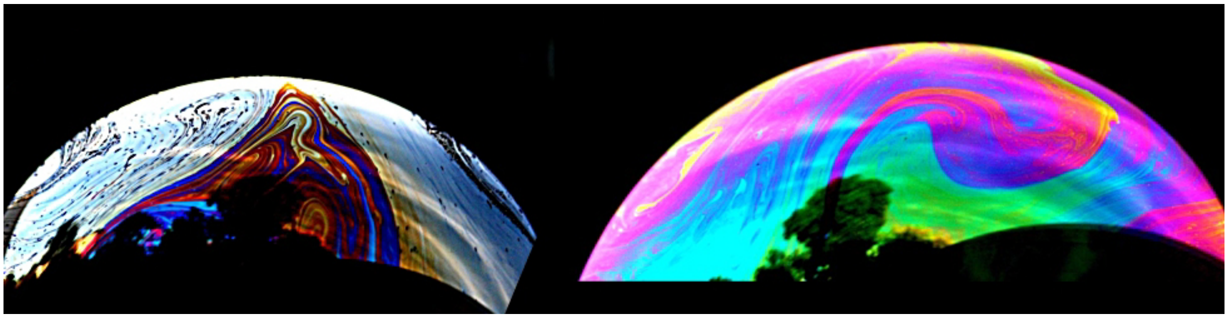
normal mode
    possible standing wave pattern for a standing wave on a string

overtone
    frequency that produces standing waves and is higher than the fundamental frequency

standing wave
    wave that can bounce back and forth through a particular region, effectively becoming stationary

Introduction
class="introduction"

Soap bubbles are blown from clear fluid into very thin films. The colors we see are not due to any pigmentation but are the result of light interference, which enhances specific wavelengths for a given thickness of the film.



The most certain indication of a wave is interference. This wave characteristic is most prominent when the wave interacts with an object that is not large compared with the wavelength. Interference is observed for water waves, sound waves, light waves, and, in fact, all types of waves.

If you have ever looked at the reds, blues, and greens in a sunlit soap bubble and wondered how straw-colored soapy water could produce them, you have hit upon one of the many phenomena that can only be explained by the wave character of light (see [link]). The same is true for the colors seen in an oil slick or in the light reflected from a DVD disc. These and other interesting phenomena cannot be explained fully by geometric optics. In these cases, light interacts with objects and exhibits wave characteristics. The branch of optics that considers the behavior of light when it exhibits wave characteristics is called wave optics (sometimes called physical optics). It is the topic of this chapter.

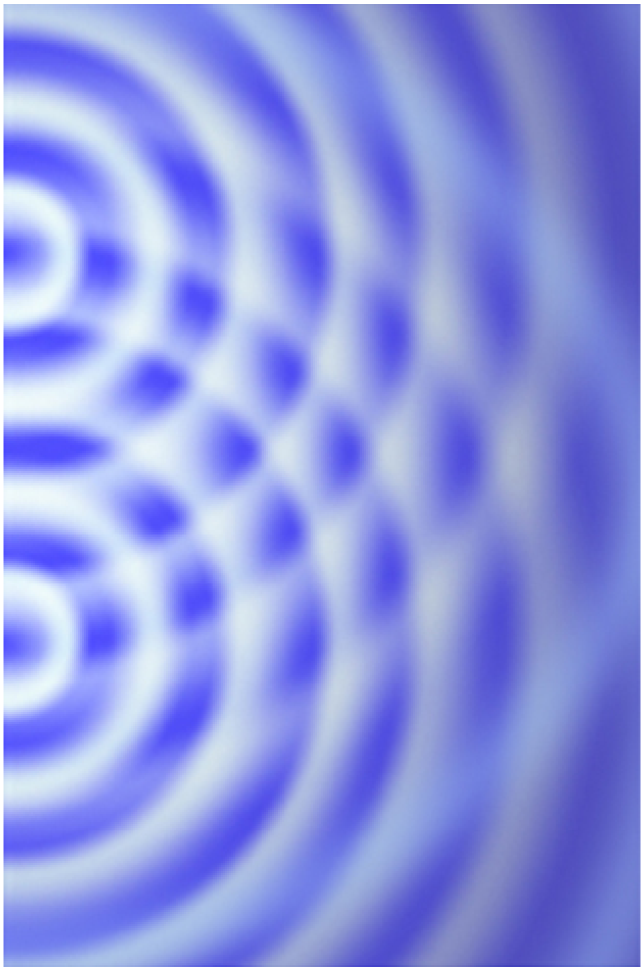
Young's Double-Slit Interference

By the end of this section, you will be able to:

- Explain the phenomenon of interference
- Define constructive and destructive interference for a double slit

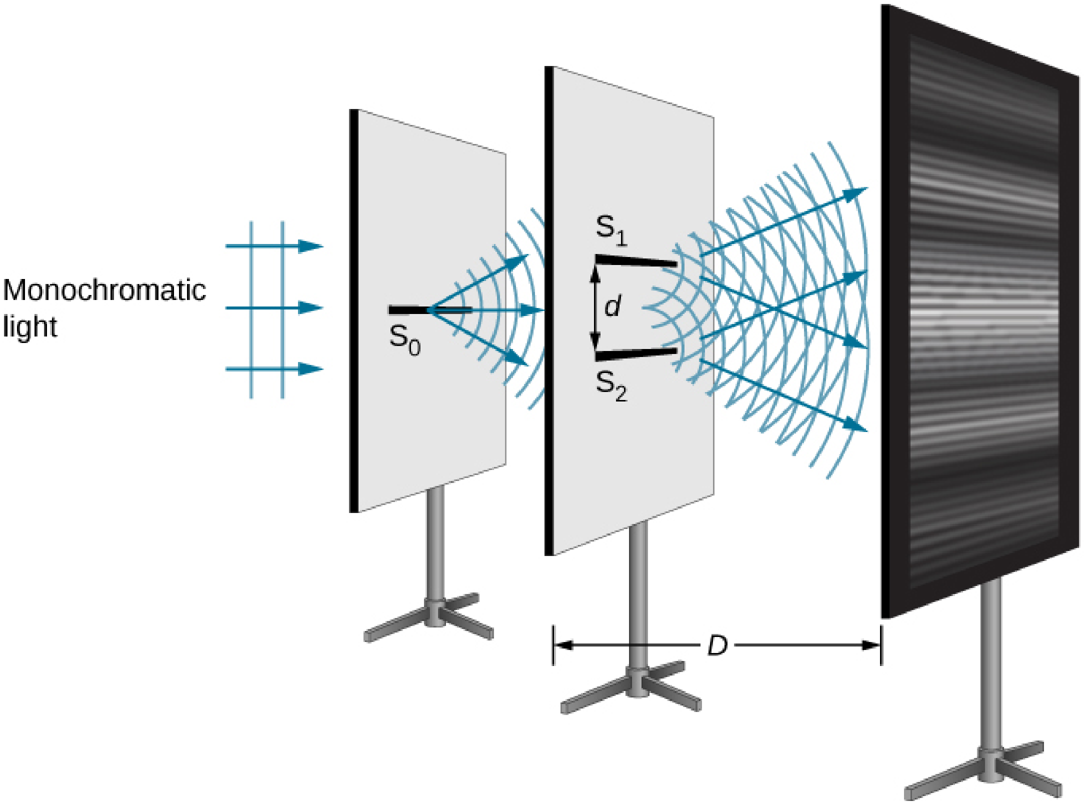
The Dutch physicist Christiaan Huygens (1629–1695) thought that light was a wave, but Isaac Newton did not. Newton thought that there were other explanations for color, and for the interference and diffraction effects that were observable at the time. Owing to Newton’s tremendous reputation, his view generally prevailed; the fact that Huygens’s principle worked was not considered direct evidence proving that light is a wave. The acceptance of the wave character of light came many years later in 1801, when the English physicist and physician Thomas Young (1773–1829) demonstrated optical interference with his now-classic double-slit experiment.

If there were not one but two sources of waves, the waves could be made to interfere, as in the case of waves on water ([link]). If light is an electromagnetic wave, it must therefore exhibit interference effects under appropriate circumstances. In Young’s experiment, sunlight was passed through a pinhole on a board. The emerging beam fell on two pinholes on a second board. The light emanating from the two pinholes then fell on a screen where a pattern of bright and dark spots was observed. This pattern, called fringes, can only be explained through interference, a wave phenomenon.



Photograph of an interference pattern produced by circular water waves in a ripple tank. Two thin plungers are vibrated up and down in phase at the surface of the water. Circular water waves are produced by and emanate from each plunger. The points where the water is calm (corresponding to destructive interference) are clearly visible.

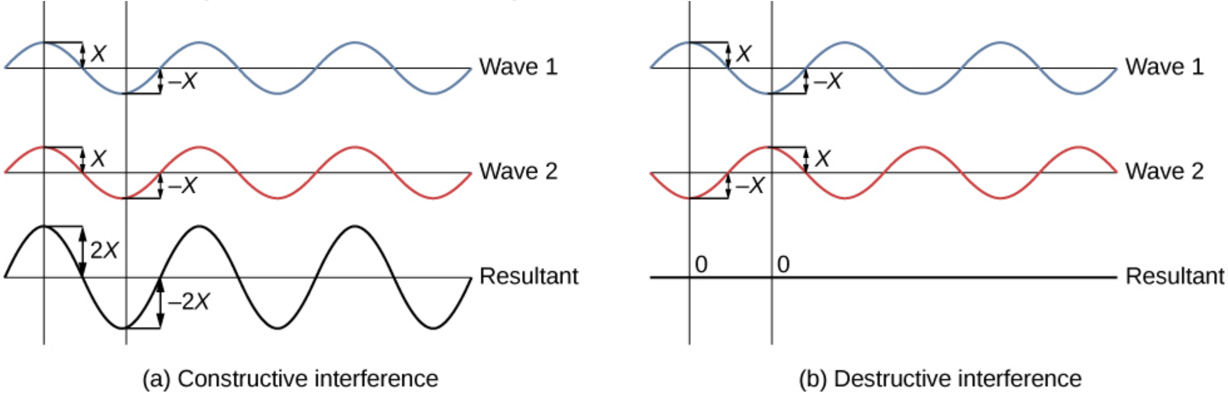
We can analyze double-slit interference with the help of [link], which depicts an apparatus analogous to Young’s. Light from a monochromatic source falls on a slit S_0. The light emanating from S_0 is incident on two other slits S_1 and S_2 that are equidistant from S_0. A pattern of *interference fringes* on the screen is then produced by the light emanating from S_1 and S_2. All slits are assumed to be so narrow that they can be considered secondary point sources for Huygens' wavelets (The Nature of Light). Slits S_1 and S_2 are a distance d apart ($d \le 1\ \mathrm{mm}$), and the distance between the screen and the slits is $D(\approx 1\ \mathrm{m})$, which is much greater than d.



The double-slit interference experiment using monochromatic light and narrow slits. Fringes produced by interfering Huygens wavelets from slits S_1 and S_2 are observed on the screen.

Since S_0 is assumed to be a point source of monochromatic light, the secondary Huygens wavelets leaving S_1 and S_2 always maintain a constant phase difference (zero in this case because S_1 and S_2 are equidistant from S_0) and have the same frequency. The sources S_1 and S_2 are then said to be coherent. By **coherent waves**, we mean the waves are in phase or have a definite phase relationship. The term **incoherent** means the waves have random phase relationships, which would be the case if S_1 and S_2 were illuminated by two independent light sources, rather than a single source S_0. Two independent light sources (which may be two separate areas within the same lamp or the Sun) would generally not emit their light in unison, that is, not coherently. Also, because S_1 and S_2 are the same distance from S_0, the amplitudes of the two Huygens wavelets are equal.

Young used sunlight, where each wavelength forms its own pattern, making the effect more difficult to see. In the following discussion, we illustrate the double-slit experiment with **monochromatic** light (single λ) to clarify the effect. [link] shows the pure constructive and destructive interference of two waves having the same wavelength and amplitude.

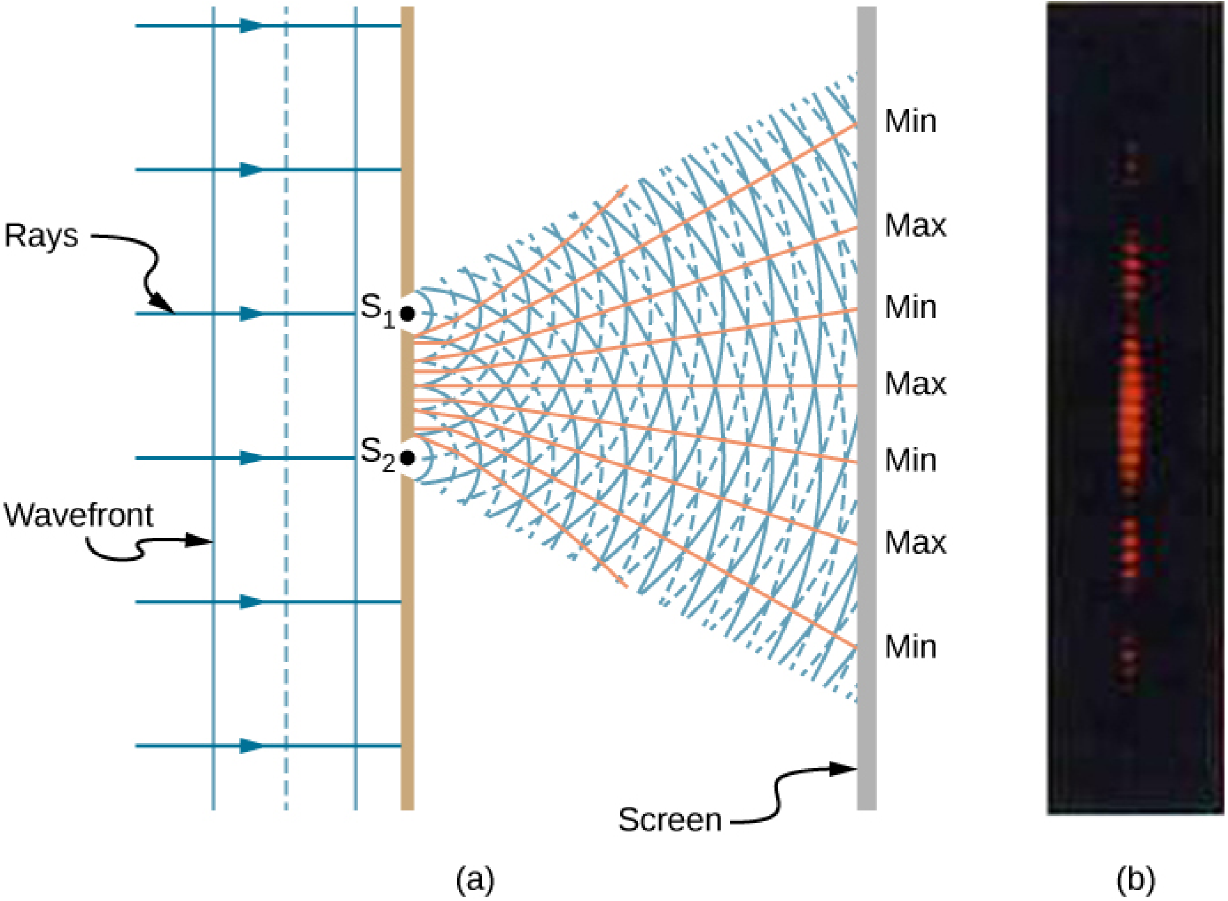


(a) Constructive interference

(b) Destructive interference

The amplitudes of waves add. (a) Pure constructive interference is obtained when identical waves are in phase. (b) Pure destructive interference occurs when identical waves are exactly out of phase, or shifted by half a wavelength.

When light passes through narrow slits, the slits act as sources of coherent waves and light spreads out as semicircular waves, as shown in [link](a). Pure *constructive interference* occurs where the waves are crest to crest or trough to trough. Pure *destructive interference* occurs where they are crest to trough. The light must fall on a screen and be scattered into our eyes for us to see the pattern. An analogous pattern for water waves is shown in [link]. Note that regions of constructive and destructive interference move out from the slits at well-defined angles to the original beam. These angles depend on wavelength and the distance between the slits, as we shall see below.



Double slits produce two coherent sources of waves that interfere. (a) Light spreads out (diffracts) from each slit, because the slits are narrow. These waves overlap and interfere constructively (bright lines) and destructively (dark regions). We can only see this if the light falls onto a screen and is scattered into our eyes. (b) When light that has passed through double slits falls on a screen, we see a pattern such as this.

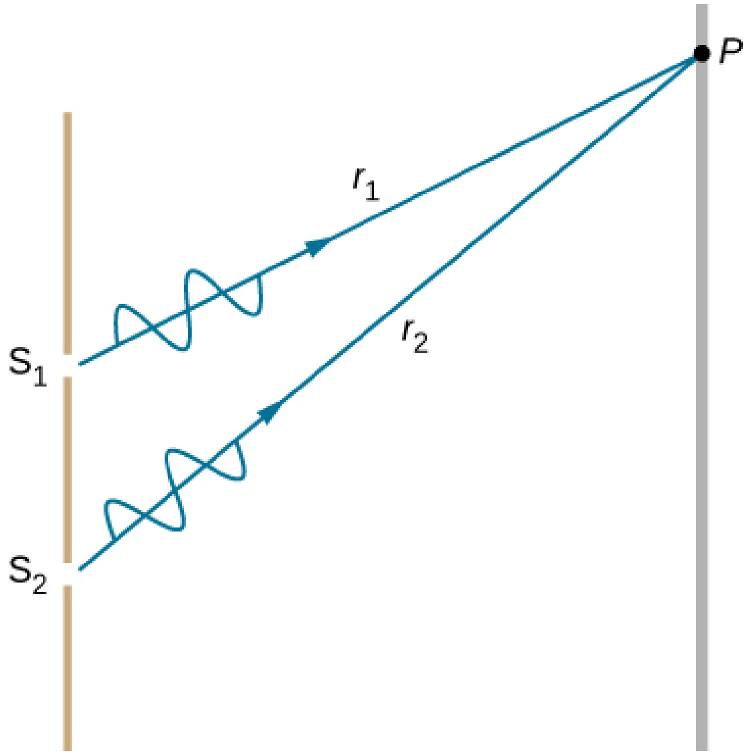
To understand the double-slit interference pattern, consider how two waves travel from the slits to the screen ([link]). Each slit is a different distance from a given point on the screen. Thus, different numbers of wavelengths fit into each path. Waves start out from the slits in phase (crest to crest), but they may end up out of phase (crest to trough) at the screen if the paths differ in length by half a wavelength, interfering destructively. If the paths differ by a whole wavelength, then the waves arrive in phase (crest to crest) at the screen, interfering constructively. More generally, if the path length difference Δl between the two waves is any half-integral number of wavelengths [(1 / 2)λ, (3 / 2)λ, (5 / 2)λ, etc.], then destructive interference occurs. Similarly, if the path length difference is any integral number of wavelengths (λ, 2λ, 3λ, etc.), then constructive interference occurs. These conditions can be expressed as equations:

Equation:

$$\Delta l = m\lambda, \quad \text{for } m = 0, \pm1, \pm2, \pm3 \ \ldots \ \text{(constructive interference)}$$

Equation:

$$\Delta l = (m + \frac{1}{2})\lambda, \quad \text{for } m = 0, \pm1, \pm2, \pm3 \ \ldots \ \text{(destructive interference)}$$



Waves follow different paths from the slits to a common point P on a screen. Destructive interference occurs where one path is a half wavelength longer than the other—the waves start in phase but arrive out of phase. Constructive interference occurs where one path is a whole wavelength longer than the other—the waves start out and arrive in phase.

Summary

- Young’s double-slit experiment gave definitive proof of the wave character of light.
- An interference pattern is obtained by the superposition of light from two slits.

Conceptual Questions

Exercise:

Problem:

Young’s double-slit experiment breaks a single light beam into two sources. Would the same pattern be obtained for two independent sources of light, such as the headlights of a distant car? Explain.

Solution:

No. Two independent light sources do not have coherent phase.

Exercise:

Problem:

Is it possible to create a experimental setup in which there is only destructive interference? Explain.

Exercise:

Problem:

Why won’t two small sodium lamps, held close together, produce an interference pattern on a distant screen? What if the sodium lamps were replaced by two laser pointers held close together?

Solution:

Because both the sodium lamps are not coherent pairs of light sources. Two lasers operating independently are also not coherent so no interference pattern results.

Glossary

coherent waves
waves are in phase or have a definite phase relationship

incoherent
waves have random phase relationships

monochromatic
light composed of one wavelength only

Mathematics of Interference

By the end of this section, you will be able to:

- Determine the angles for bright and dark fringes for double slit interference
- Calculate the positions of bright fringes on a screen

[link](a) shows how to determine the path length difference Δl for waves traveling from two slits to a common point on a screen. If the screen is a large distance away compared with the distance between the slits, then the angle θ between the path and a line from the slits to the screen [part (b)] is nearly the same for each path. In other words, r_1 and r_2 are essentially parallel. The lengths of r_1 and r_2 differ by Δl, as indicated by the two dashed lines in the figure. Simple trigonometry shows

Equation:

$$\Delta l = d \sin \theta$$

where d is the distance between the slits. Combining this result with [link], we obtain constructive interference for a double slit when the path length difference is an integral multiple of the wavelength, or

Equation:

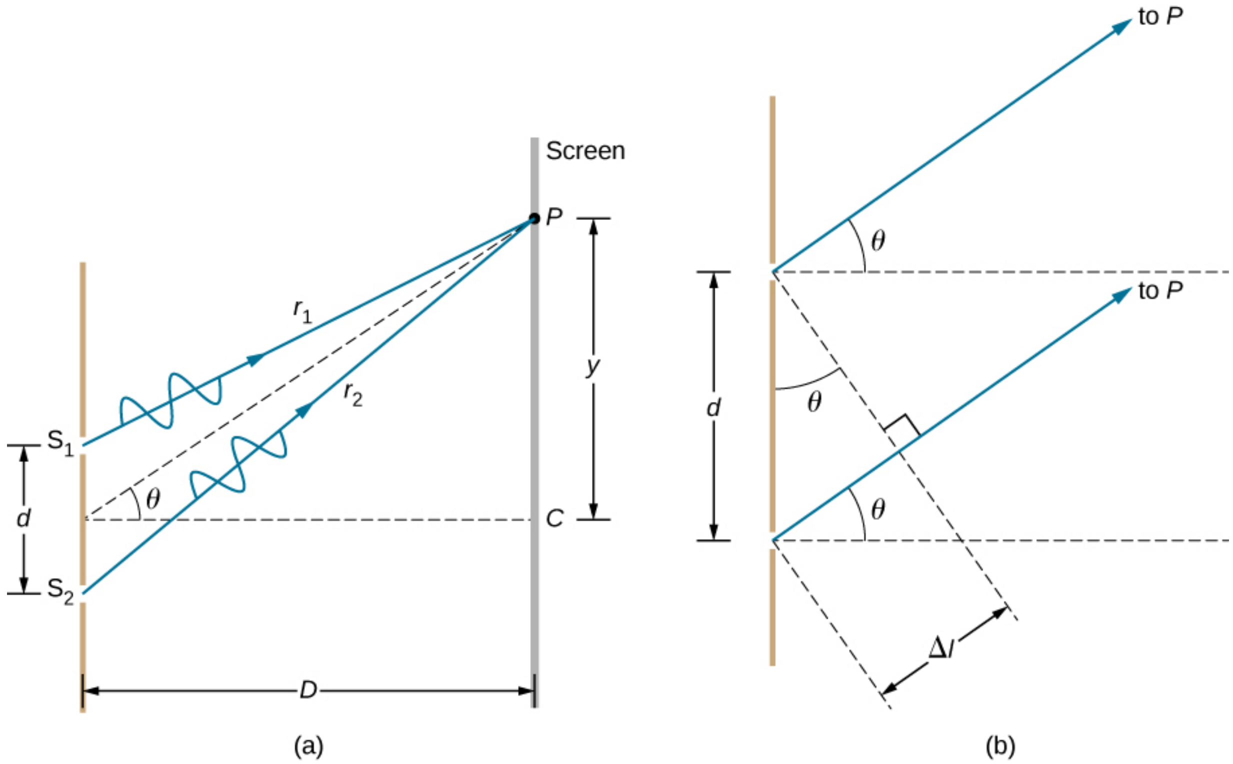
$$d \sin \theta = m\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3, \ldots \text{ (constructive interference).}$$

Similarly, to obtain destructive interference for a double slit, the path length difference must be a half-integral multiple of the wavelength, or

Equation:

$$d \sin \theta = (m + \frac{1}{2})\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3, \ldots \text{ (destructive interference)}$$

where λ is the wavelength of the light, d is the distance between slits, and θ is the angle from the original direction of the beam as discussed above. We call m the **order** of the interference. For example, $m = 4$ is fourth-order interference.



(a) To reach P, the light waves from S_1 and S_2 must travel different distances. (b) The path difference between the two rays is Δl.

The equations for double-slit interference imply that a series of bright and dark lines are formed. For vertical slits, the light spreads out horizontally on either side of the incident beam into a pattern called interference **fringes** ([link]). The closer the slits are, the more the bright fringes spread apart. We can see this by examining the equation

$d \sin\theta = m\lambda,\ \text{for}\ m = 0, \pm1, \pm2, \pm3\ldots$. For fixed λ and m, the smaller d is, the larger θ must be, since $\sin\theta = m\lambda/d$. This is consistent with our contention that wave effects are most noticeable when the object the wave encounters (here, slits a distance d apart) is small. Small d gives large θ, hence, a large effect.

Referring back to part (a) of the figure, θ is typically small enough that $\sin\theta \approx \tan\theta \approx y_m/D$, where y_m is the distance from the central maximum to the mth bright fringe and D is the distance between the slit and the screen. [link] may then be written as

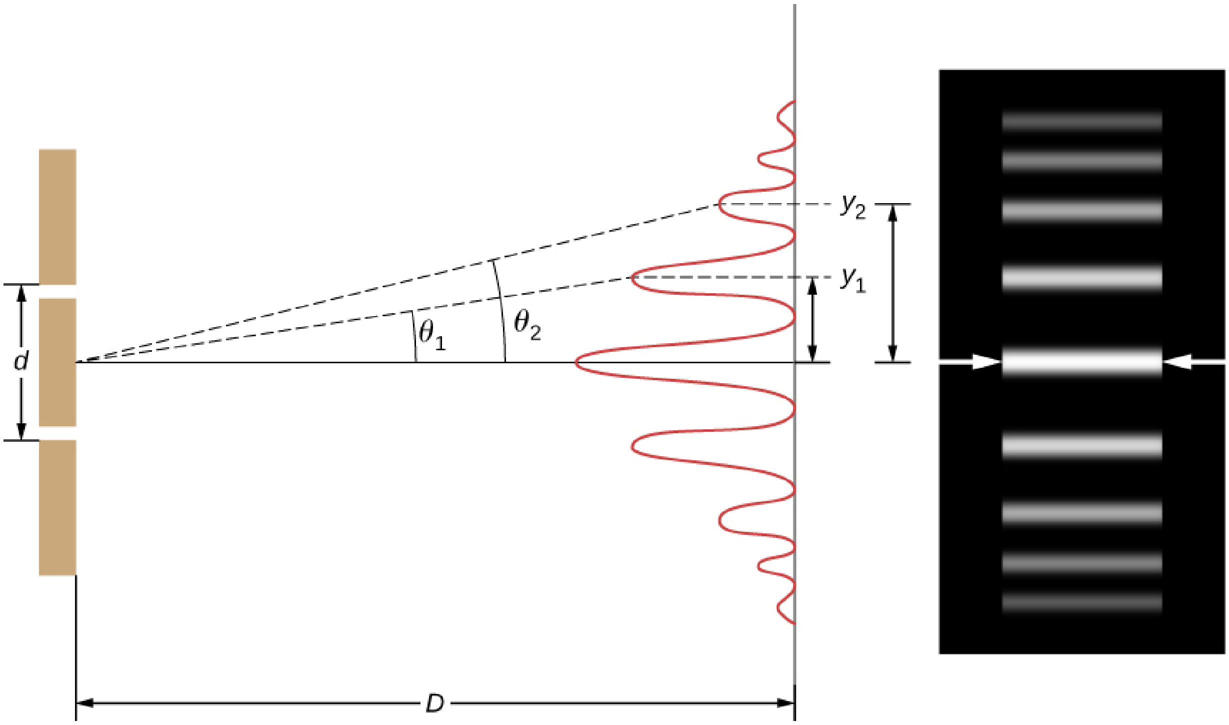
Equation:

$$d\frac{y_m}{D} = m\lambda$$

or

Equation:

$$y_m = \frac{m\lambda D}{d}.$$



The interference pattern for a double slit has an intensity that falls off with angle. The image shows multiple bright and dark lines, or fringes, formed by light passing through a double slit.

Example:
Finding a Wavelength from an Interference Pattern

Suppose you pass light from a He-Ne laser through two slits separated by 0.0100 mm and find that the third bright line on a screen is formed at an angle of 10.95° relative to the incident beam. What is the wavelength of the light?

Strategy

The phenomenon is two-slit interference as illustrated in [link] and the third bright line is due to third-order constructive interference, which means that $m = 3$. We are given $d = 0.0100\ \text{mm}$ and $\theta = 10.95^\circ$. The wavelength can thus be found using the equation $d\sin\theta = m\lambda$ for constructive interference.

Solution

Solving $d\sin\theta = m\lambda$ for the wavelength λ gives

Equation:

$$\lambda = \frac{d\sin\theta}{m}.$$

Substituting known values yields

Equation:

$$\lambda = \frac{(0.0100\,\text{mm})(\sin 10.95^\circ)}{3} = 6.33 \times 10^{-4}\,\text{mm} = 633\,\text{nm}.$$

Significance

To three digits, this is the wavelength of light emitted by the common He-Ne laser. Not by coincidence, this red color is similar to that emitted by neon lights. More important, however, is the fact that interference patterns can be used to measure wavelength. Young did this for visible wavelengths. This analytical techinque is still widely used to measure electromagnetic spectra. For a given order, the angle for constructive interference increases with λ, so that spectra (measurements of intensity versus wavelength) can be obtained.

Example:

Calculating the Highest Order Possible

Interference patterns do not have an infinite number of lines, since there is a limit to how big *m* can be. What is the highest-order constructive interference possible with the system described in the preceding example?

Strategy

The equation $d\sin\theta = m\lambda$ (for $m = 0, \pm1, \pm2, \pm3\ldots$) describes constructive interference from two slits. For fixed values of d and λ, the larger m is, the larger $\sin\theta$ is. However, the maximum value that $\sin\theta$ can have is 1, for an angle of 90°. (Larger angles imply that light goes backward and does not reach the screen at all.) Let us find what value of m corresponds to this maximum diffraction angle.

Solution

Solving the equation $d\sin\theta = m\lambda$ for m gives

Equation:

$$m = \frac{d\sin\theta}{\lambda}.$$

Taking $\sin\theta = 1$ and substituting the values of d and λ from the preceding example gives

Equation:

$$m = \frac{(0.0100\,\text{mm})(1)}{633\,\text{nm}} \approx 15.8.$$

Therefore, the largest integer m can be is 15, or $m = 15$.

Significance

The number of fringes depends on the wavelength and slit separation. The number of fringes is very large for large slit separations. However, recall (see The Propagation of Light and the introduction for this chapter) that wave interference is only prominent when the wave interacts with objects that are not large compared to the wavelength. Therefore, if the slit separation and the sizes of the slits become much greater than the wavelength, the intensity pattern of light on the screen changes, so there are simply two bright lines cast by the slits, as expected, when light behaves like rays. We also note that the fringes get fainter farther away from the center. Consequently, not all 15 fringes may be observable.

Note:

Exercise:

Problem:

Check Your Understanding In the system used in the preceding examples, at what angles are the first and the second bright fringes formed?

Solution:

3.63° and 7.27°, respectively

Summary

- In double-slit diffraction, constructive interference occurs when $d \sin \theta = m\lambda$ (for $m = 0, \pm 1, \pm 2, \pm 3 \ldots$), where d is the distance between the slits, θ is the angle relative to the incident direction, and m is the order of the interference.
- Destructive interference occurs when $d \sin \theta = (m + \frac{1}{2})\lambda$ for $m \ = \ 0, \pm 1, \pm 2, \pm 3, \ldots$.

Conceptual Questions

Exercise:

Problem:

Suppose you use the same double slit to perform Young's double-slit experiment in air and then repeat the experiment in water. Do the angles to the same parts of the interference pattern get larger or smaller? Does the color of the light change? Explain.

Exercise:

Problem:

Why is monochromatic light used in the double slit experiment? What would happen if white light were used?

Solution:

Monochromatic sources produce fringes at angles according to $d \sin \theta = m\lambda$. With white light, each constituent wavelength will produce fringes at its own set of angles, blending into the fringes of adjacent wavelengths. This results in rainbow patterns.

Problems

Exercise:

Problem:

At what angle is the first-order maximum for 450-nm wavelength blue light falling on double slits separated by 0.0500 mm?

Exercise:

Problem:

Calculate the angle for the third-order maximum of 580-nm wavelength yellow light falling on double slits separated by 0.100 mm.

Solution:

0.997°

Exercise:

Problem:

What is the separation between two slits for which 610-nm orange light has its first maximum at an angle of 30.0°?

Exercise:

Problem:

Find the distance between two slits that produces the first minimum for 410-nm violet light at an angle of 45.0°.

Solution:

$0.290\ \mu\text{m}$

Exercise:

Problem:

Calculate the wavelength of light that has its third minimum at an angle of 30.0° when falling on double slits separated by $3.00\ \mu\text{m}$. Explicitly show how you follow the steps from the Problem-Solving Strategy: Wave Optics, located at the end of the chapter.

Exercise:

Problem:

What is the wavelength of light falling on double slits separated by $2.00\ \mu\text{m}$ if the third-order maximum is at an angle of 60.0°?

Solution:

$5.77\ \times\ 10^{-7}\,\text{m} = 577\,\text{nm}$

Exercise:

Problem:

At what angle is the fourth-order maximum for the situation in the preceding problem?

Exercise:

Problem:

What is the highest-order maximum for 400-nm light falling on double slits separated by $25.0\ \mu\text{m}$?

Solution:

62.5; since *m* must be an integer, the highest order is then $m = 62$.

Exercise:

Problem:

Find the largest wavelength of light falling on double slits separated by $1.20\ \mu\text{m}$ for which there is a first-order maximum. Is this in the visible part of the spectrum?

Exercise:

Problem:

What is the smallest separation between two slits that will produce a second-order maximum for 720-nm red light?

Solution:

$1.44\ \mu\text{m}$

Exercise:

Problem:

(a) What is the smallest separation between two slits that will produce a second-order maximum for any visible light? (b) For all visible light?

Exercise:

Problem:

(a) If the first-order maximum for monochromatic light falling on a double slit is at an angle of 10.0°, at what angle is the second-order maximum? (b) What is the angle of the first minimum? (c) What is the highest-order maximum possible here?

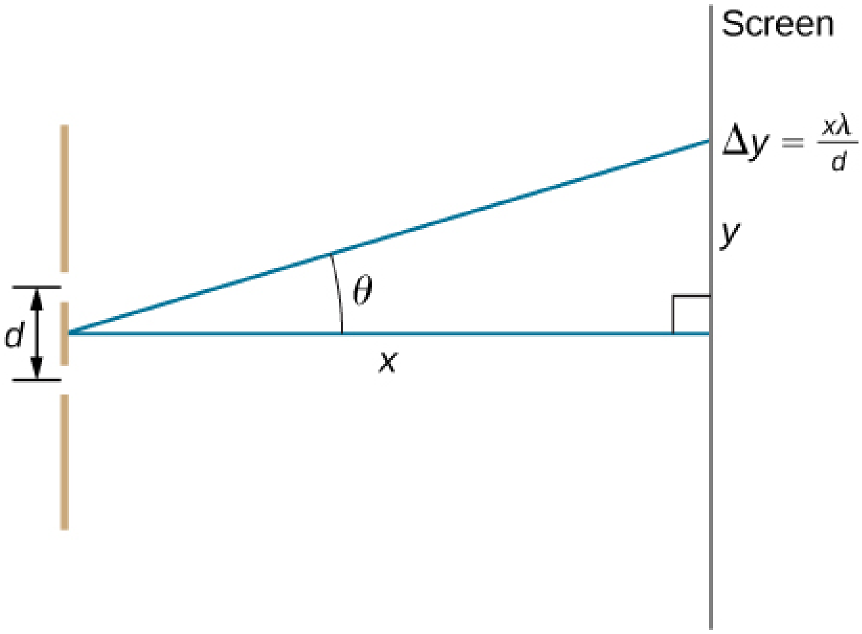
Solution:

a. 20.3°; b. 4.98°; c. 5.76, the highest order is $m = 5$.

Exercise:

Problem:

Shown below is a double slit located a distance *x* from a screen, with the distance from the center of the screen given by *y*. When the distance *d* between the slits is relatively large, numerous bright spots appear, called fringes. Show that, for small angles (where $\sin\theta \approx \theta$, with θ in radians), the distance between fringes is given by $\Delta y = x\lambda/d$



Exercise:

Problem:

Using the result of the preceding problem, (a) calculate the distance between fringes for 633-nm light falling on double slits separated by 0.0800 mm, located 3.00 m from a screen. (b) What would be the distance between fringes if the entire apparatus were submersed in water, whose index of refraction is 1.33?

Solution:

a. 2.37 cm; b. 1.78 cm

Exercise:

Problem:

Using the result of the problem two problems prior, find the wavelength of light that produces fringes 7.50 mm apart on a screen 2.00 m from double slits separated by 0.120 mm.

Exercise:

Problem:

In a double-slit experiment, the fifth maximum is 2.8 cm from the central maximum on a screen that is 1.5 m away from the slits. If the slits are 0.15 mm apart, what is the wavelength of the light being used?

Solution:

560 nm

Exercise:

Problem:

The source in Young’s experiment emits at two wavelengths. On the viewing screen, the fourth maximum for one wavelength is located at the same spot as the fifth maximum for the other wavelength. What is the ratio of the two wavelengths?

Exercise:

Problem:

If 500-nm and 650-nm light illuminates two slits that are separated by 0.50 mm, how far apart are the second-order maxima for these two wavelengths on a screen 2.0 m away?

Solution:

1.2 mm

Exercise:

Problem:

Red light of wavelength of 700 nm falls on a double slit separated by 400 nm. (a) At what angle is the first-order maximum in the diffraction pattern? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

fringes
bright and dark patterns of interference

order
integer *m* used in the equations for constructive and destructive interference for a double slit

Multiple-Slit Interference

By the end of this section, you will be able to:

- Describe the locations and intensities of secondary maxima for multiple-slit interference

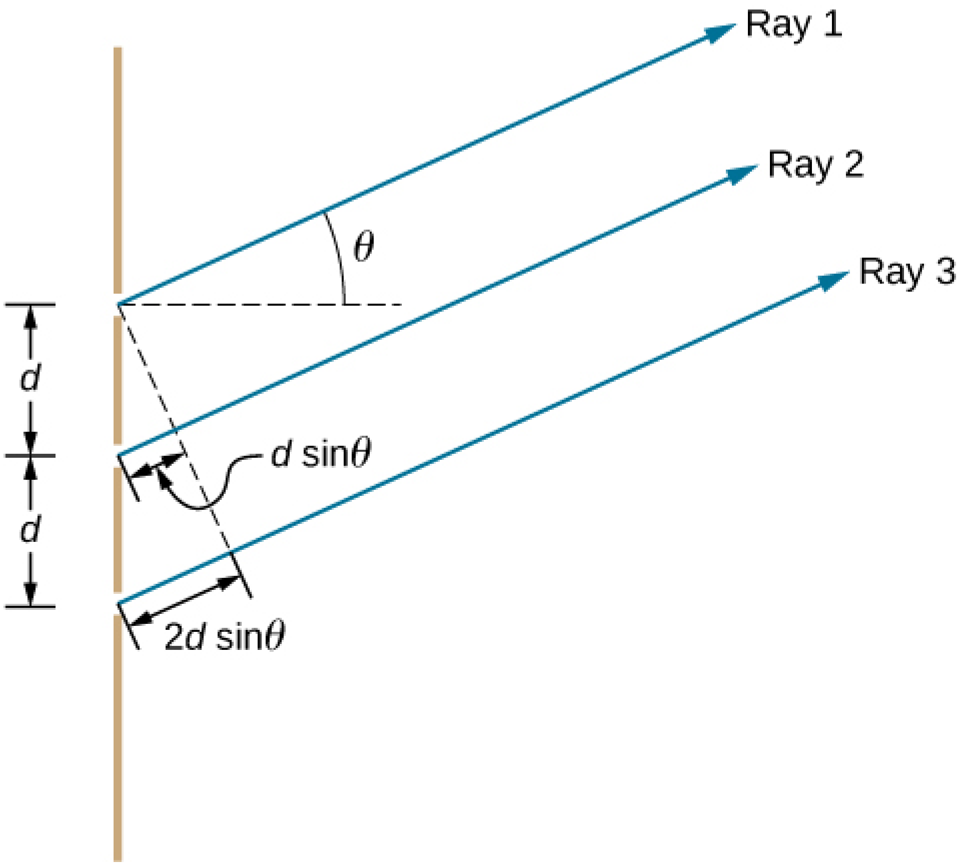
Analyzing the interference of light passing through two slits lays out the theoretical framework of interference and gives us a historical insight into Thomas Young’s experiments. However, much of the modern-day application of slit interference uses not just two slits but many, approaching infinity for practical purposes. The key optical element is called a diffraction grating, an important tool in optical analysis, which we discuss in detail in Diffraction. Here, we start the analysis of multiple-slit interference by taking the results from our analysis of the double slit ($N=2$) and extending it to configurations with three, four, and much larger numbers of slits.

[link] shows the simplest case of multiple-slit interference, with three slits, or $N=3$. The spacing between slits is d, and the path length difference between adjacent slits is $d\sin\theta$, same as the case for the double slit. What is new is that the path length difference for the first and the third slits is $2d\sin\theta$. The condition for constructive interference is the same as for the double slit, that is

Equation:

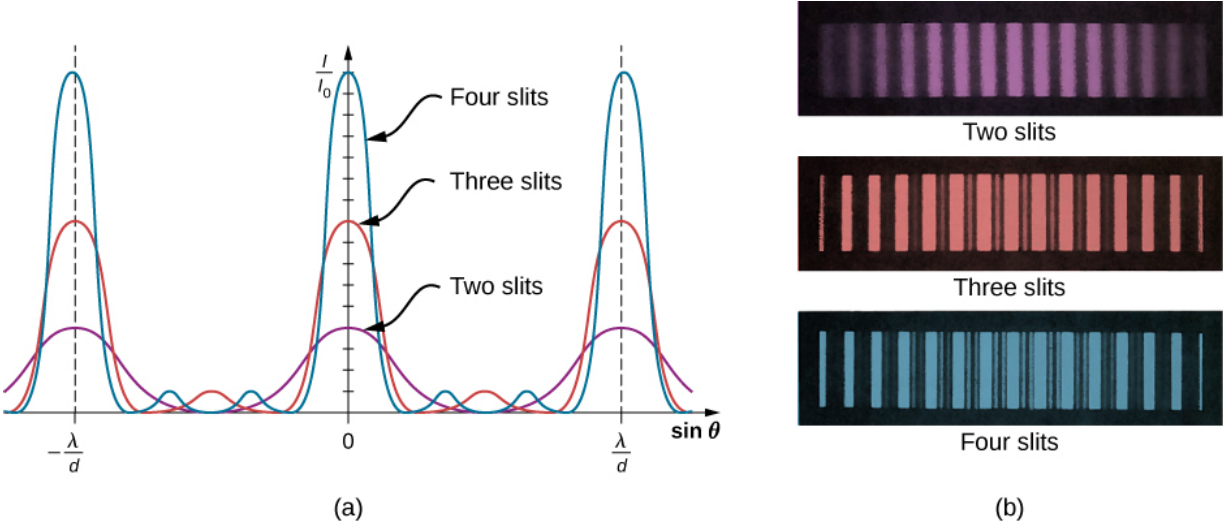
$$d\sin\theta=m\lambda.$$

When this condition is met, $2d\sin\theta$ is automatically a multiple of λ, so all three rays combine constructively, and the bright fringes that occur here are called **principal maxima**. But what happens when the path length difference between adjacent slits is only $\lambda/2$? We can think of the first and second rays as interfering destructively, but the third ray remains unaltered. Instead of obtaining a dark fringe, or a minimum, as we did for the double slit, we see a **secondary maximum** with intensity lower than the principal maxima.



Interference with three slits. Different pairs of emerging rays can combine constructively or destructively at the same time, leading to secondary maxima.

In general, for N slits, these secondary maxima occur whenever an unpaired ray is present that does not go away due to destructive interference. This occurs at $(N-2)$ evenly spaced positions between the principal maxima. The amplitude of the electromagnetic wave is correspondingly diminished to $1/N$ of the wave at the principal maxima, and the light intensity, being proportional to the square of the wave amplitude, is diminished to $1/N^2$ of the intensity compared to the principal maxima. As [link] shows, a dark fringe is located between every maximum (principal or secondary). As N grows larger and the number of bright and dark fringes increase, the widths of the maxima become narrower due to the closely located neighboring dark fringes. Because the total amount of light energy remains unaltered, narrower maxima require that each maximum reaches a correspondingly higher intensity.



Interference fringe patterns for two, three and four slits. As the number of slits increases, more secondary maxima appear, but the principal maxima become brighter and narrower. (a) Graph and (b) photographs of fringe patterns.

Summary

- Interference from multiple slits ($N > 2$) produces principal as well as secondary maxima.
- As the number of slits is increased, the intensity of the principal maxima increases and the width decreases.

Problems

Exercise:

Problem:

Ten narrow slits are equally spaced 0.25 mm apart and illuminated with yellow light of wavelength 580 nm. (a) What are the angular positions of the third and fourth principal maxima? (b) What is the separation of these maxima on a screen 2.0 m from the slits?

Solution:

a. $0.40^\circ, 0.53^\circ$; b. $4.6 \times 10^{-3}\,\text{m}$

Exercise:

Problem:

The width of bright fringes can be calculated as the separation between the two adjacent dark fringes on either side. Find the angular widths of the third- and fourth-order bright fringes from the preceding problem.

Exercise:

Problem:

For a three-slit interference pattern, find the ratio of the peak intensities of a secondary maximum to a principal maximum.

Solution:

1:9

Exercise:

Problem:

What is the angular width of the central fringe of the interference pattern of (a) 20 slits separated by $d = 2.0 \times 10^{-3}\text{mm}$? (b) 50 slits with the same separation? Assume that $\lambda = 600\,\text{nm}$.

Glossary

principal maximum	brightest interference fringes seen with multiple slits
secondary maximum	bright interference fringes of intensity lower than the principal maxima

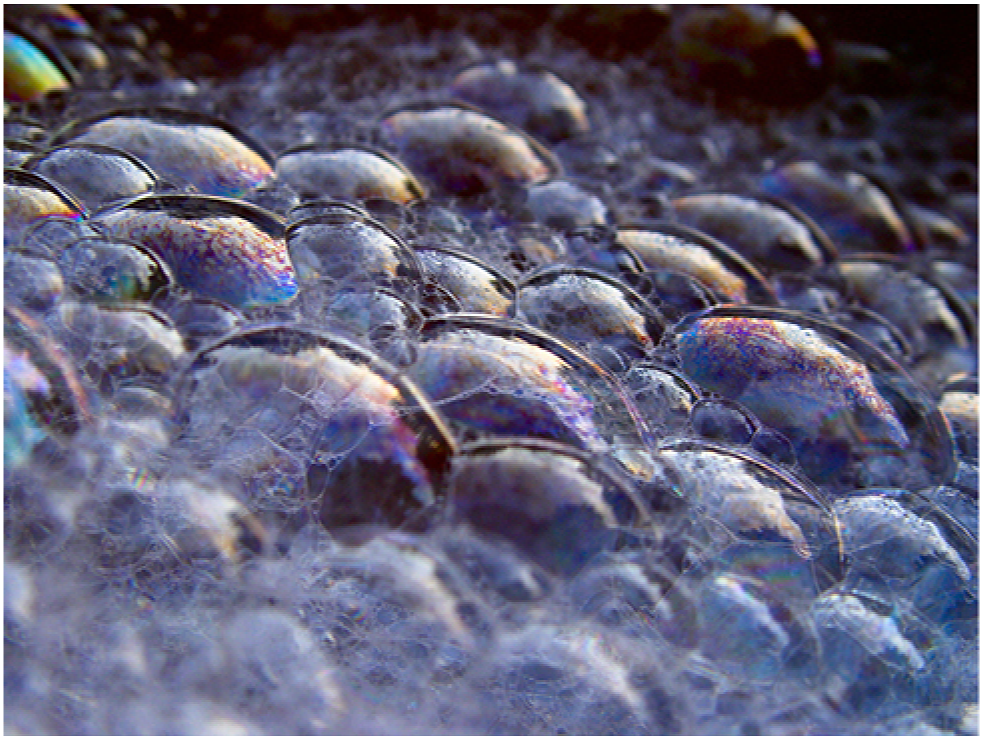
Interference in Thin Films

By the end of this section, you will be able to:

- Describe the phase changes that occur upon reflection
- Describe fringes established by reflected rays of a common source
- Explain the appearance of colors in thin films

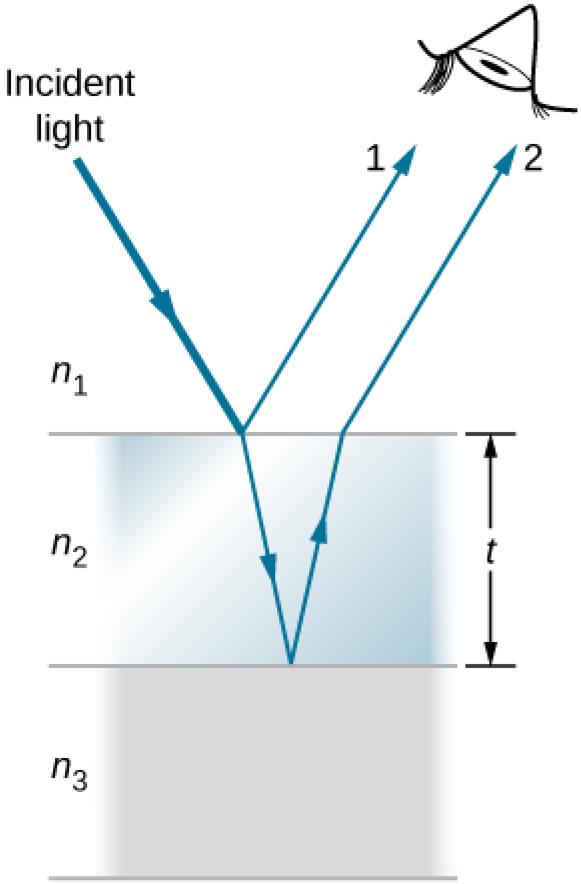
The bright colors seen in an oil slick floating on water or in a sunlit soap bubble are caused by interference. The brightest colors are those that interfere constructively. This interference is between light reflected from different surfaces of a thin film; thus, the effect is known as **thin-film interference**.

As we noted before, interference effects are most prominent when light interacts with something having a size similar to its wavelength. A thin film is one having a thickness t smaller than a few times the wavelength of light, λ. Since color is associated indirectly with λ and because all interference depends in some way on the ratio of λ to the size of the object involved, we should expect to see different colors for different thicknesses of a film, as in [link].



These soap bubbles exhibit brilliant colors when exposed to sunlight. (credit: Scott Robinson)

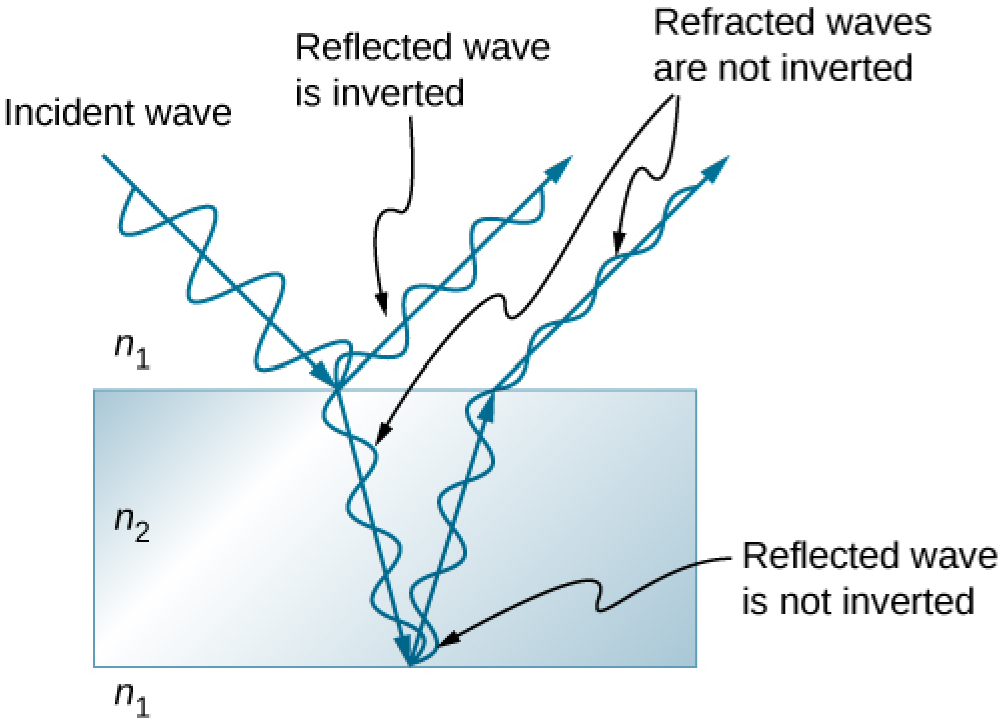
What causes thin-film interference? [link] shows how light reflected from the top and bottom surfaces of a film can interfere. Incident light is only partially reflected from the top surface of the film (ray 1). The remainder enters the film and is itself partially reflected from the bottom surface. Part of the light reflected from the bottom surface can emerge from the top of the film (ray 2) and interfere with light reflected from the top (ray 1). The ray that enters the film travels a greater distance, so it may be in or out of phase with the ray reflected from the top. However, consider for a moment, again, the bubbles in [link]. The bubbles are darkest where they are thinnest. Furthermore, if you observe a soap bubble carefully, you will note it gets dark at the point where it breaks. For very thin films, the difference in path lengths of rays 1 and 2 in [link] is negligible, so why should they interfere destructively and not constructively? The answer is that a phase change can occur upon reflection, as discussed next.



Light striking a thin film is partially reflected (ray 1) and partially refracted at the top surface. The refracted ray is partially reflected at the bottom surface and emerges as ray 2. These rays interfere in a way that depends on the thickness of the film and the indices of refraction of the various media.

Changes in Phase due to Reflection

We saw earlier (Waves) that reflection of mechanical waves can involve a 180° phase change. For example, a traveling wave on a string is inverted (i.e., a 180° phase change) upon reflection at a boundary to which a heavier string is tied. However, if the second string is lighter (or more precisely, of a lower linear density), no inversion occurs. Light waves produce the same effect, but the deciding parameter for light is the index of refraction. Light waves undergo a 180° or π radians phase change upon reflection at an interface beyond which is a medium of higher index of refraction. No phase change takes place when reflecting from a medium of lower refractive index ([link]). Because of the periodic nature of waves, this phase change or inversion is equivalent to $\pm\lambda/2$ in distance travelled, or path length. Both the path length and refractive indices are important factors in thin-film interference.



Reflection at an interface for light traveling from a medium with index of refraction n_1 to a medium with index of refraction n_2, $n_1 < n_2$, causes the phase of the wave to change by π radians.

If the film in [link] is a **soap bubble** (essentially water with air on both sides), then a phase shift of $\lambda/2$ occurs for ray 1 but not for ray 2. Thus, when the film is very thin and the path length difference between the two rays is negligible, they are exactly out of phase, and destructive interference occurs at all wavelengths. Thus, the soap bubble is dark here. The thickness of the film relative to the wavelength of light is the other crucial factor in thin-film interference. Ray 2 in [link] travels a greater distance than ray 1. For light incident perpendicular to the surface, ray 2 travels a distance approximately 2*t* farther than ray 1. When this distance is an integral or half-integral multiple of the wavelength in the medium $(\lambda_n = \lambda/n$, where λ is the wavelength in vacuum and *n* is the index of refraction), constructive or destructive interference occurs, depending also on whether there is a phase change in either ray.

Example:

Calculating the Thickness of a Nonreflective Lens Coating

Sophisticated cameras use a series of several lenses. Light can reflect from the surfaces of these various lenses and degrade image clarity. To limit these reflections, lenses are coated with a thin layer of magnesium fluoride, which causes destructive thin-film interference. What is the thinnest this film can be, if its index of refraction is 1.38 and it is designed to limit the reflection of 550-nm light, normally the most intense visible wavelength? Assume the index of refraction of the glass is 1.52.

Strategy

Refer to [link] and use $n_1 = 1.00$ for air, $n_2 = 1.38$, and $n_3 = 1.52$. Both ray 1 and ray 2 have a $\lambda/2$ shift upon reflection. Thus, to obtain destructive interference, ray 2 needs to travel a half wavelength farther than ray 1. For rays incident perpendicularly, the path length difference is 2*t*.

Solution

To obtain destructive interference here,

Equation:

$$2t = \frac{\lambda_{n2}}{2}$$

where λ_{n2} is the wavelength in the film and is given by $\lambda_{n2} = \lambda / n_2$. Thus,

Equation:

$$2t = \frac{\lambda / n_2}{2}.$$

Solving for *t* and entering known values yields

Equation:

$$t = \frac{\lambda / n_2}{4} = \frac{(500\,\text{nm})/1.38}{4} = 90.6\,\text{nm}.$$

Significance

Films such as the one in this example are most effective in producing destructive interference when the thinnest layer is used, since light over a broader range of incident angles is reduced in intensity. These films are called nonreflective coatings; this is only an approximately correct description, though, since other wavelengths are only partially cancelled. Nonreflective coatings are also used in car windows and sunglasses.

Combining Path Length Difference with Phase Change

Thin-film interference is most constructive or most destructive when the path length difference for the two rays is an integral or half-integral wavelength. That is, for rays incident perpendicularly,

Equation:

$$2t = \lambda_n, 2\lambda_n, 3\lambda_n, \ldots \text{ or } 2t = \lambda_n/2, 3\lambda_n/2, 5\lambda_n/2, \ldots.$$

To know whether interference is constructive or destructive, you must also determine if there is a phase change upon reflection. Thin-film interference thus depends on film thickness, the wavelength of light, and the refractive indices. For white light incident on a film that varies in thickness, you can observe rainbow colors of constructive interference for various wavelengths as the thickness varies.

Example:

Soap Bubbles

(a) What are the three smallest thicknesses of a soap bubble that produce constructive interference for red light with a wavelength of 650 nm? The index of refraction of soap is taken to be the same as that of water. (b) What three smallest thicknesses give destructive interference?

Strategy

Use [link] to visualize the bubble, which acts as a thin film between two layers of air. Thus $n_1 = n_3 = 1.00$ for air, and $n_2 = 1.333$ for soap (equivalent to water). There is a $\lambda/2$ shift for ray 1 reflected from the top surface of the bubble and no shift for ray 2 reflected from the bottom surface. To get constructive interference, then, the path length difference (2*t*) must be a half-integral multiple of the wavelength—the first three being $\lambda_n/2, 3\lambda_n/2$, and $5\lambda_n/2$. To get destructive interference, the path length difference must be an integral multiple of the wavelength—the first three being 0, λ_n, and $2\lambda_n$.

Solution

a. Constructive interference occurs here when

Equation:

$$2t_{\text{c}} = \frac{\lambda_n}{2}, \frac{3\lambda_n}{2}, \frac{5\lambda_n}{2}, \ldots .$$

Thus, the smallest constructive thickness t_{c} is

Equation:

$$t_{\text{c}} = \frac{\lambda_n}{4} = \frac{\lambda/n}{4} = \frac{(650\,\text{nm})/1.333}{4} = 122\,\text{nm}.$$

The next thickness that gives constructive interference is $t'_{\text{c}} = 3\lambda_n/4$, so that

Equation:

$$t'_{\text{c}} = 366\,\text{nm}.$$

Finally, the third thickness producing constructive interference is $t'_{\text{c}} = 5\lambda_n/4$, so that

Equation:

$$t'_{\text{c}} = 610\,\text{nm}.$$

b. For destructive interference, the path length difference here is an integral multiple of the wavelength. The first occurs for zero thickness, since there is a phase change at the top surface, that is,

Equation:

$$t_{\text{d}} = 0,$$

the very thin (or negligibly thin) case discussed above. The first non-zero thickness producing destructive interference is

Equation:

$$2t'_{\text{d}} = \lambda_n.$$

Substituting known values gives

Equation:

$$t'_{\text{d}} = \frac{\lambda}{2} = \frac{\lambda/n}{2} = \frac{(650\,\text{nm})/1.333}{2} = 244\,\text{nm}.$$

Finally, the third destructive thickness is $2t''_{\text{d}} = 2\lambda_n$, so that

Equation:

$$t''_{\text{d}} = \lambda_n = \frac{\lambda}{n} = \frac{650\,\text{nm}}{1.333} = 488\,\text{nm}.$$

Significance

If the bubble were illuminated with pure red light, we would see bright and dark bands at very uniform increases in thickness. First would be a dark band at 0 thickness, then bright at 122 nm thickness, then dark at 244 nm, bright at 366 nm, dark at 488 nm, and bright at 610 nm. If the bubble varied smoothly in thickness, like a smooth wedge, then the bands would be evenly spaced.

Note:
Exercise:

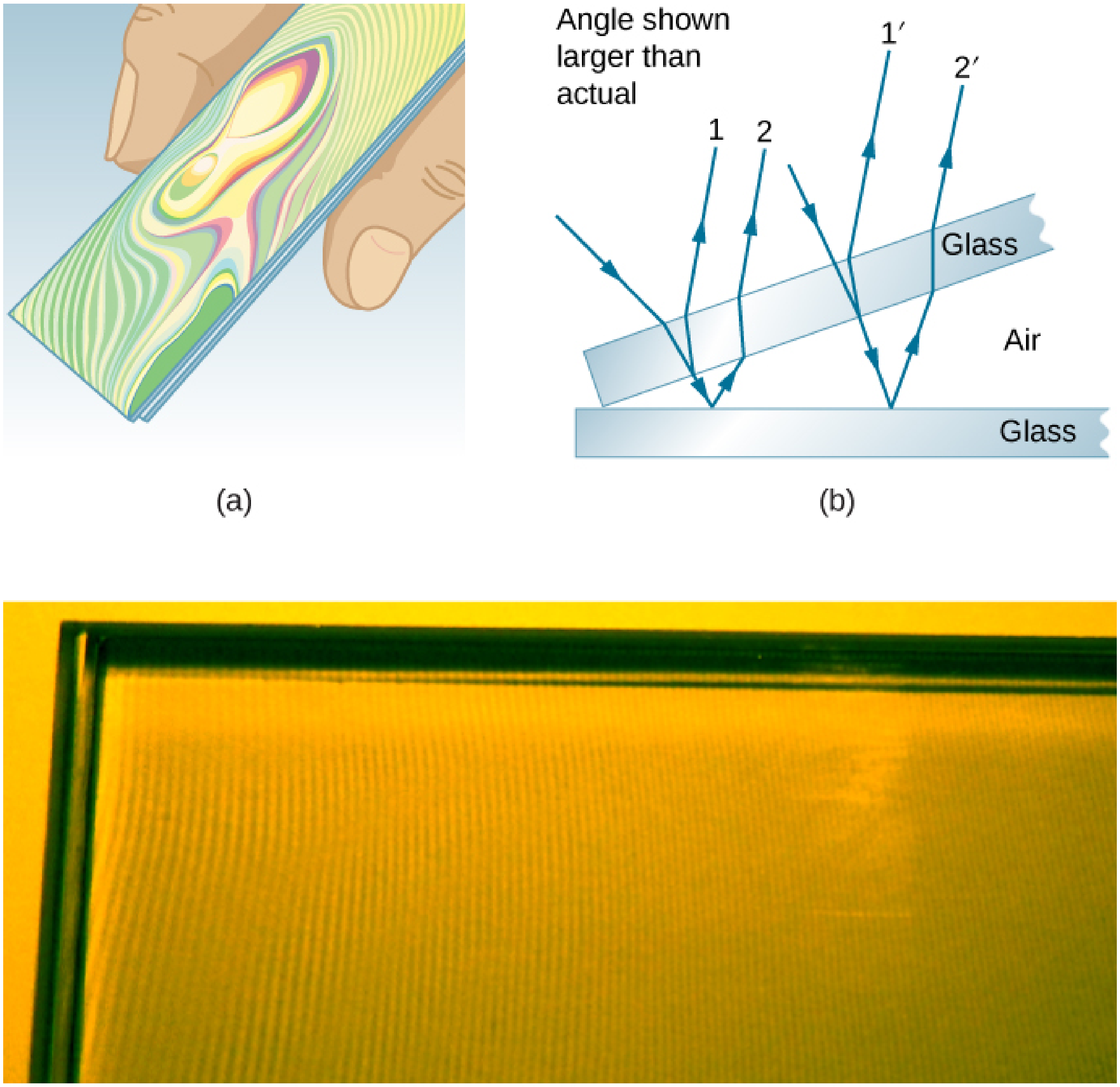
Problem:

Check Your Understanding Going further with [link], what are the next two thicknesses of soap bubble that would lead to (a) constructive interference, and (b) destructive interference?

Solution:

a. 853 nm, 1097 nm; b. 731 nm, 975 nm

Another example of thin-film interference can be seen when microscope slides are separated (see [link]). The slides are very flat, so that the wedge of air between them increases in thickness very uniformly. A phase change occurs at the second surface but not the first, so a dark band forms where the slides touch. The rainbow colors of constructive interference repeat, going from violet to red again and again as the distance between the slides increases. As the layer of air increases, the bands become more difficult to see, because slight changes in incident angle have greater effects on path length differences. If monochromatic light instead of white light is used, then bright and dark bands are obtained rather than repeating rainbow colors.



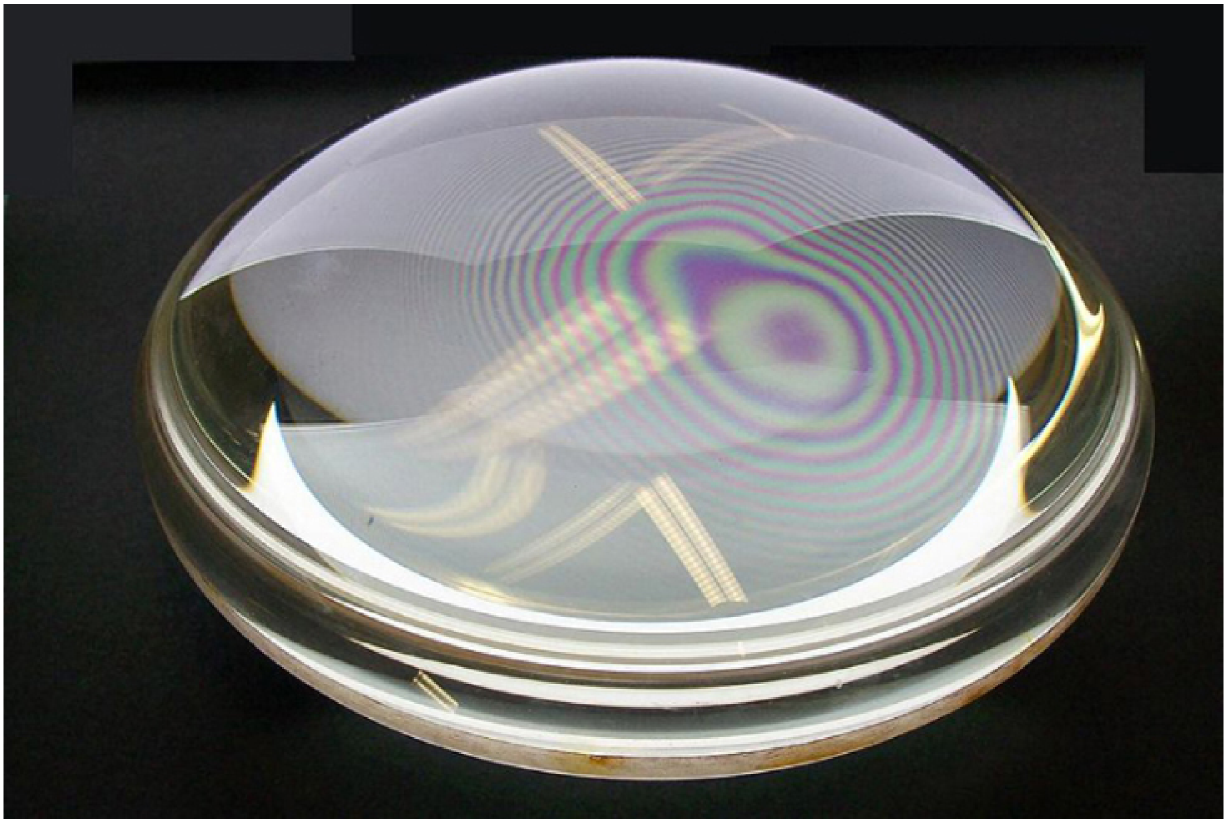
(a)

(b)

(c)

(a) The rainbow-color bands are produced by thin-film interference in the air between the two glass slides. (b) Schematic of the paths taken by rays in the wedge of air between the slides. (c) If the air wedge is illuminated with monochromatic light, bright and dark bands are obtained rather than repeating rainbow colors.

An important application of thin-film interference is found in the manufacturing of optical instruments. A lens or mirror can be compared with a master as it is being ground, allowing it to be shaped to an accuracy of less than a wavelength over its entire surface. [link] illustrates the phenomenon called **Newton's rings**, which occurs when the plane surfaces of two lenses are placed together. (The circular bands are called Newton's rings because Isaac Newton described them and their use in detail. Newton did not discover them; Robert Hooke did, and Newton did not believe they were due to the wave character of light.) Each successive ring of a given color indicates an increase of only half a wavelength in the distance between the lens and the blank, so that great precision can be obtained. Once the lens is perfect, no rings appear.



"Newton's rings" interference fringes are produced when two plano-convex lenses are placed together with their plane surfaces in contact. The rings are created by interference between the light reflected off the two surfaces as a result of a slight gap between them, indicating that these surfaces are not precisely plane but are slightly convex. (credit: Ulf Seifert)

Thin-film interference has many other applications, both in nature and in manufacturing. The wings of certain moths and butterflies have nearly iridescent colors due to thin-film interference. In addition to pigmentation, the wing’s color is affected greatly by constructive interference of certain wavelengths reflected from its film-coated surface. Some car manufacturers offer special paint jobs that use thin-film interference to produce colors that change with angle. This expensive option is based on variation of thin-film path length differences with angle. Security features on credit cards, banknotes, driving licenses, and similar items prone to forgery use thin-film interference, diffraction gratings, or holograms. As early as 1998, Australia led the way with dollar bills printed on polymer with a diffraction grating security feature, making the currency difficult to forge. Other countries, such as Canada, New Zealand, and Taiwan, are using similar technologies, while US currency includes a thin-film interference effect.

Summary

- When light reflects from a medium having an index of refraction greater than that of the medium in which it is traveling, a 180° phase change (or a $\lambda/2$ shift) occurs.
- Thin-film interference occurs between the light reflected from the top and bottom surfaces of a film. In addition to the path length difference, there can be a phase change.

Conceptual Questions

Exercise:

Problem:

What effect does increasing the wedge angle have on the spacing of interference fringes? If the wedge angle is too large, fringes are not observed. Why?

Exercise:

Problem:

How is the difference in paths taken by two originally in-phase light waves related to whether they interfere constructively or destructively? How can this be affected by reflection? By refraction?

Solution:

Differing path lengths result in different phases at destination resulting in constructive or destructive interference accordingly. Reflection can cause a 180° phase change, which also affects how waves interfere. Refraction into another medium changes the wavelength inside that medium such that a wave can emerge from the medium with a different phase compared to another wave that travelled the same distance in a different medium.

Exercise:

Problem:

Is there a phase change in the light reflected from either surface of a contact lens floating on a person's tear layer? The index of refraction of the lens is about 1.5, and its top surface is dry.

Exercise:

Problem:

In placing a sample on a microscope slide, a glass cover is placed over a water drop on the glass slide. Light incident from above can reflect from the top and bottom of the glass cover and from the glass slide below the water drop. At which surfaces will there be a phase change in the reflected light?

Solution:

Phase changes occur upon reflection at the top of glass cover and the top of glass slide only.

Exercise:

Problem:

Answer the above question if the fluid between the two pieces of crown glass is carbon disulfide.

Exercise:

Problem:

While contemplating the food value of a slice of ham, you notice a rainbow of color reflected from its moist surface. Explain its origin.

Solution:

The surface of the ham being moist means there is a thin layer of fluid, resulting in thin-film interference. Because the exact thickness of the film varies across the piece of ham, which is illuminated by white light, different wavelengths produce bright fringes at different locations, resulting in rainbow colors.

Exercise:

Problem:

An inventor notices that a soap bubble is dark at its thinnest and realizes that destructive interference is taking place for all wavelengths. How could she use this knowledge to make a nonreflective coating for lenses that is effective at all wavelengths? That is, what limits would there be on the index of refraction and thickness of the coating? How might this be impractical?

Exercise:

Problem:

A nonreflective coating like the one described in [link] works ideally for a single wavelength and for perpendicular incidence. What happens for other wavelengths and other incident directions? Be specific.

Solution:

Other wavelengths will not generally satisfy $t = \frac{\lambda/n}{4}$ for the same value of t so reflections will result in completely destructive interference. For an incidence angle θ, the path length inside the coating will be increased by a factor $1/\cos\theta$ so the new condition for destructive interference becomes $\frac{t}{\cos\theta} = \frac{\lambda/n}{4}$.

Exercise:

Problem:

Why is it much more difficult to see interference fringes for light reflected from a thick piece of glass than from a thin film? Would it be easier if monochromatic light were used?

Problems

Exercise:

Problem:

A soap bubble is 100 nm thick and illuminated by white light incident perpendicular to its surface. What wavelength and color of visible light is most constructively reflected, assuming the same index of refraction as water?

Solution:

532 nm (green)

Exercise:

Problem:

An oil slick on water is 120 nm thick and illuminated by white light incident perpendicular to its surface. What color does the oil appear (what is the most constructively reflected wavelength), given its index of refraction is 1.40?

Exercise:

Problem:

Calculate the minimum thickness of an oil slick on water that appears blue when illuminated by white light perpendicular to its surface. Take the blue wavelength to be 470 nm and the index of refraction of oil to be 1.40.

Solution:

$8.39 \times 10^{-8}\,\text{m} = 83.9\,\text{nm}$

Exercise:

Problem:

Find the minimum thickness of a soap bubble that appears red when illuminated by white light perpendicular to its surface. Take the wavelength to be 680 nm, and assume the same index of refraction as water.

Exercise:

Problem:

A film of soapy water ($n = 1.33$) on top of a plastic cutting board has a thickness of 233 nm. What color is most strongly reflected if it is illuminated perpendicular to its surface?

Solution:

620 nm (orange)

Exercise:

Problem:

What are the three smallest non-zero thicknesses of soapy water ($n = 1.33$) on Plexiglas if it appears green (constructively reflecting 520-nm light) when illuminated perpendicularly by white light?

Exercise:

Problem:

Suppose you have a lens system that is to be used primarily for 700-nm red light. What is the second thinnest coating of fluorite (magnesium fluoride) that would be nonreflective for this wavelength?

Solution:

380 nm

Exercise:

Problem:

(a) As a soap bubble thins it becomes dark, because the path length difference becomes small compared with the wavelength of light and there is a phase shift at the top surface. If it becomes dark when the path length difference is less than one-fourth the wavelength, what is the thickest the bubble can be and appear dark at all visible wavelengths? Assume the same index of refraction as water. (b) Discuss the fragility of the film considering the thickness found.

Exercise:

Problem:

To save money on making military aircraft invisible to radar, an inventor decides to coat them with a nonreflective material having an index of refraction of 1.20, which is between that of air and the surface of the plane. This, he reasons, should be much cheaper than designing Stealth bombers. (a) What thickness should the coating be to inhibit the reflection of 4.00-cm wavelength radar? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. Assuming *n* for the plane is greater than 1.20, then there are two phase changes: 0.833 cm. b. It is too thick, and the plane would be too heavy. c. It is unreasonable to think the layer of material could be any thickness when used on a real aircraft.

Glossary

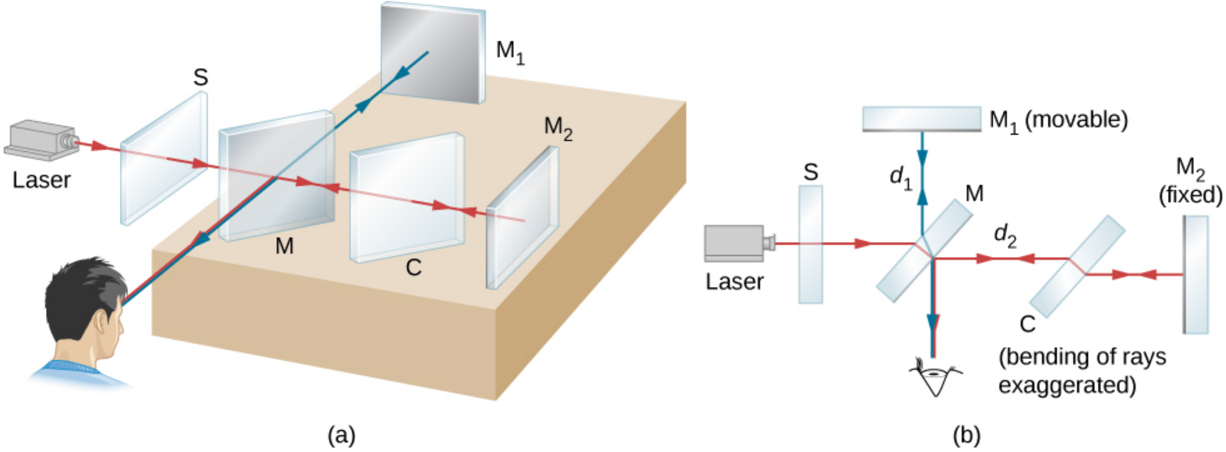
Newton’s rings	circular interference pattern created by interference between the light reflected off two surfaces as a result of a slight gap between them
thin-film interference	interference between light reflected from different surfaces of a thin film

The Michelson Interferometer

By the end of this section, you will be able to:

- Explain changes in fringes observed with a Michelson interferometer caused by mirror movements
- Explain changes in fringes observed with a Michelson interferometer caused by changes in medium

The Michelson **interferometer** (invented by the American physicist Albert A. Michelson, 1852–1931) is a precision instrument that produces interference fringes by splitting a light beam into two parts and then recombining them after they have traveled different optical paths. [link] depicts the interferometer and the path of a light beam from a single point on the extended source S, which is a ground-glass plate that diffuses the light from a monochromatic lamp of wavelength λ_0. The beam strikes the half-silvered mirror M, where half of it is reflected to the side and half passes through the mirror. The reflected light travels to the movable plane mirror M_1, where it is reflected back through M to the observer. The transmitted half of the original beam is reflected back by the stationary mirror M_2 and then toward the observer by M.



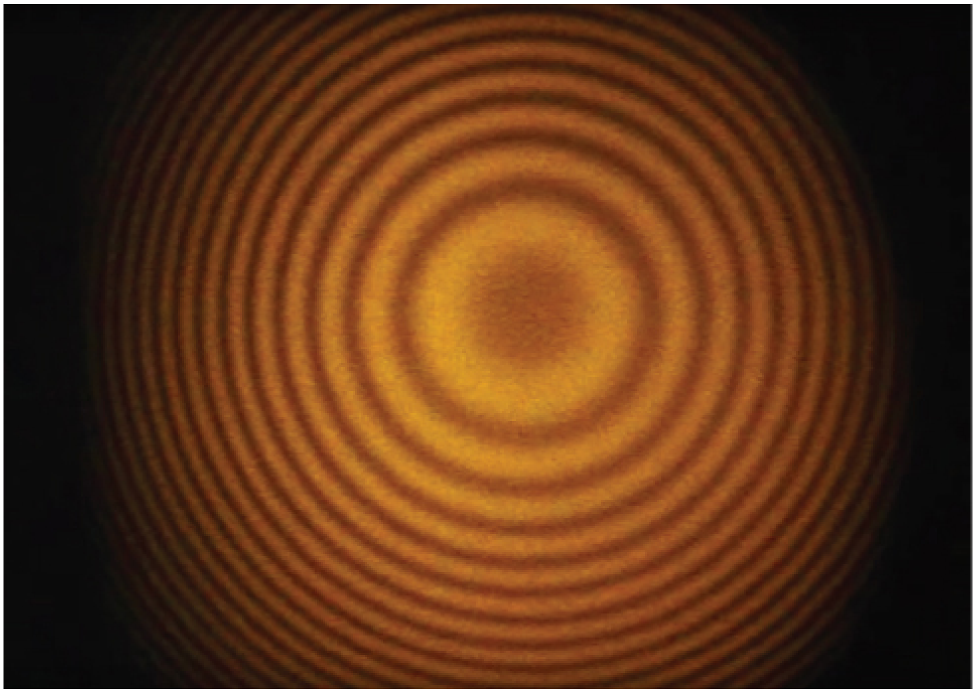
(a) The Michelson interferometer. The extended light source is a ground-glass plate that diffuses the light from a laser. (b) A planar view of the interferometer.

Because both beams originate from the same point on the source, they are coherent and therefore interfere. Notice from the figure that one beam passes through M three times and the other only once. To ensure that both beams traverse the same thickness of glass, a compensator plate C of transparent glass is placed in the arm containing M_2. This plate is a duplicate of M (without the silvering) and is usually cut from the same piece of glass used to produce M. With the compensator in place, any phase difference between the two beams is due solely to the difference in the distances they travel.

The path difference of the two beams when they recombine is $2d_1 - 2d_2$, where d_1 is the distance between M and M_1, and d_2 is the distance between M and M_2. Suppose this path difference is an integer number of wavelengths $m\lambda_0$. Then, constructive interference occurs and a bright image of the point on the source is seen at the observer. Now the light from any other point on the source whose two beams have this same path difference also undergoes constructive interference and produces a bright image. The collection of these point images is a bright fringe corresponding to a path difference of $m\lambda_0$ ([link]). When M_1 is moved a distance $\Delta d = \lambda_0/2$, this path difference changes by λ_0, and each fringe moves to the position previously occupied by an adjacent fringe. Consequently, by counting the number of fringes *m* passing a given point as M_1 is moved, an observer can measure minute displacements that are accurate to a fraction of a wavelength, as shown by the relation

Equation:

$$\Delta d = m\frac{\lambda_0}{2}.$$



Fringes produced with a Michelson interferometer.
(credit: “SILLAGESvideos”/YouTube)

Example:

Precise Distance Measurements by Michelson Interferometer

A red laser light of wavelength 630 nm is used in a Michelson interferometer. While keeping the mirror M_1 fixed, mirror M_2 is moved. The fringes are found to move past a fixed cross-hair in the viewer. Find the distance the mirror M_2 is moved for a single fringe to move past the reference line.

Strategy

Refer to [link] for the geometry. We use the result of the Michelson interferometer interference condition to find the distance moved, Δd.

Solution

For a 630-nm red laser light, and for each fringe crossing $(m = 1)$, the distance traveled by M_2 if you keep M_1 fixed is

Equation:

$$\Delta d = m\frac{\lambda_0}{2} = 1 \times \frac{630\ \text{nm}}{2} = 315\ \text{nm} = 0.315\ \mu\text{m}.$$

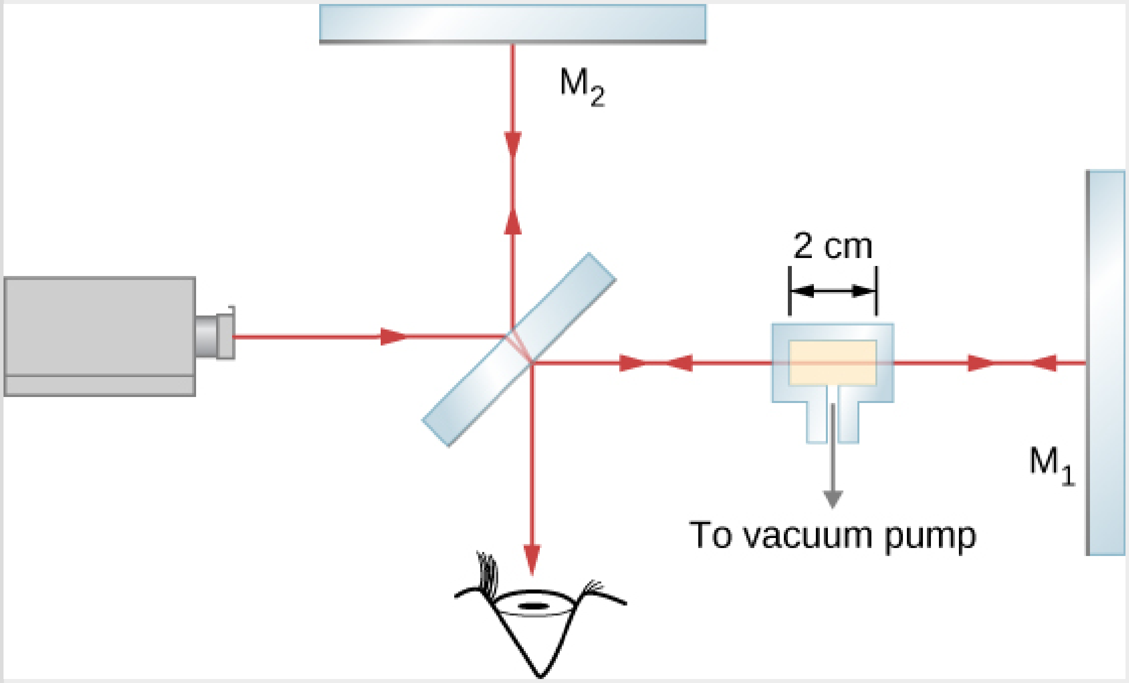
Significance

An important application of this measurement is the definition of the standard meter. As mentioned in Units and Measurement, the length of the standard meter was once defined as the mirror displacement in a Michelson interferometer corresponding to 1,650,763.73 wavelengths of the particular fringe of krypton-86 in a gas discharge tube.

Example:

Measuring the Refractive Index of a Gas

In one arm of a Michelson interferometer, a glass chamber is placed with attachments for evacuating the inside and putting gases in it. The space inside the container is 2 cm wide. Initially, the container is empty. As gas is slowly let into the chamber, you observe that dark fringes move past a reference line in the field of observation. By the time the chamber is filled to the desired pressure, you have counted 122 fringes move past the reference line. The wavelength of the light used is 632.8 nm. What is the refractive index of this gas?



Strategy

The $m = 122$ fringes observed compose the difference between the number of wavelengths that fit within the empty chamber (vacuum) and the number of wavelengths that fit within the same chamber when it is gas-filled. The wavelength in the filled chamber is shorter by a factor of *n*, the index of refraction.

Solution

The ray travels a distance $t = 2\ \text{cm}$ to the right through the glass chamber and another distance *t* to the left upon reflection. The total travel is $L = 2t$. When empty, the number of wavelengths that fit in this chamber is

Equation:

$$N_0 = \frac{L}{\lambda_0} = \frac{2t}{\lambda_0}$$

where $\lambda_0 = 632.8\ \text{nm}$ is the wavelength in vacuum of the light used. In any other medium, the wavelength is $\lambda = \lambda_0/n$ and the number of wavelengths that fit in the gas-filled chamber is

Equation:

$$N = \frac{L}{\lambda} = \frac{2t}{\lambda_0/n}.$$

The number of fringes observed in the transition is

Equation:

$$\begin{aligned} m &= N - N_0, \\ &= \tfrac{2t}{\lambda_0/n} - \tfrac{2t}{\lambda_0}, \\ &= \tfrac{2t}{\lambda_0}(n-1). \end{aligned}$$

Solving for $(n-1)$ gives

Equation:

$$n - 1 = m\left(\frac{\lambda_0}{2t}\right) = 122\left(\frac{632.8\ \times\ 10^{-9}\,\text{m}}{2(2\ \times\ 10^{-2}\,\text{m})}\right) = 0.0019$$

and $n = 1.0019$.

Significance

The indices of refraction for gases are so close to that of vacuum, that we normally consider them equal to 1. The difference between 1 and 1.0019 is so small that measuring it requires a correspondingly sensitive technique such as interferometry. We cannot, for example, hope to measure this value using techniques based simply on Snell's law.

Note:

Exercise:

Problem:

Check Your Understanding Although *m*, the number of fringes observed, is an integer, which is often regarded as having zero uncertainty, in practical terms, it is all too easy to lose track when counting fringes. In [link], if you estimate that you might have missed as many as five fringes when you reported $m = 122$ fringes, (a) is the value for the index of refraction worked out in [link] too large or too small? (b) By how much?

Solution:

a. too small; b. up to 8×10^{-5}

Note:

Problem-Solving Strategy: Wave Optics

Step 1. *Examine the situation to determine that interference is involved*. Identify whether slits, thin films, or interferometers are considered in the problem.

Step 2. *If slits are involved*, note that diffraction gratings and double slits produce very similar interference patterns, but that gratings have narrower (sharper) maxima. Single-slit patterns are characterized by a large central maximum and smaller maxima to the sides.

Step 3. *If thin-film interference or an interferometer is involved, take note of the path length difference between the two rays that interfere*. Be certain to use the wavelength in the medium involved, since it differs from the wavelength in vacuum. Note also that there is an additional $\lambda/2$ phase shift when light reflects from a medium with a greater index of refraction.
Step 4. *Identify exactly what needs to be determined in the problem (identify the unknowns)*. A written list is useful. Draw a diagram of the situation. Labeling the diagram is useful.
Step 5. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns)*.
Step 6. *Solve the appropriate equation for the quantity to be determined (the unknown) and enter the knowns*. Slits, gratings, and the Rayleigh limit involve equations.
Step 7. *For thin-film interference, you have constructive interference for a total shift that is an integral number of wavelengths. You have destructive interference for a total shift of a half-integral number of wavelengths*. Always keep in mind that crest to crest is constructive whereas crest to trough is destructive.
Step 8. *Check to see if the answer is reasonable: Does it make sense?* Angles in interference patterns cannot be greater than 90°, for example.

Summary

- When the mirror in one arm of the interferometer moves a distance of $\lambda/2$ each fringe in the interference pattern moves to the position previously occupied by the adjacent fringe.

Key Equations

Constructive	$\Delta l = m\lambda,$ for *m* = 0, ±1, ±2, ±3…

interference	
Destructive interference	$\Delta l = (m + \frac{1}{2})\lambda, \quad$ for m = 0, ±1, ±2, ±3…
Path length difference for waves from two slits to a common point on a screen	$\Delta l = d \sin \theta$
Constructive interference	$d \sin \theta = m\lambda, \quad \text{for } m = 0, \pm 1, \pm 2, \pm 3, \ldots$
Destructive interference	$d \sin \theta = (m + \frac{1}{2})\lambda, \ \text{for } m \ = \ 0, \pm 1, \pm 2, \pm 3, \ldots$
Distance from central maximum to the *m*th bright fringe	$y_m = \frac{m\lambda D}{d}$
Displacement measured by a Michelson interferometer	$\Delta d = m\frac{\lambda_0}{2}$

Conceptual Questions

Exercise:

Problem:

Describe how a Michelson interferometer can be used to measure the index of refraction of a gas (including air).

Solution:

In one arm, place a transparent chamber to be filled with the gas. See [link].

Problems

Exercise:

Problem:

A Michelson interferometer has two equal arms. A mercury light of wavelength 546 nm is used for the interferometer and stable fringes are found. One of the arms is moved by 1.5μm. How many fringes will cross the observing field?

Exercise:

Problem:

What is the distance moved by the traveling mirror of a Michelson interferometer that corresponds to 1500 fringes passing by a point of the observation screen? Assume that the interferometer is illuminated with a 606 nm spectral line of krypton-86.

Solution:

$4.55\ \times\ 10^{-4}\,\text{m}$

Exercise:

Problem:

When the traveling mirror of a Michelson interferometer is moved $2.40 \times 10^{-5}\,\text{m}$, 90 fringes pass by a point on the observation screen. What is the wavelength of the light used?

Exercise:

Problem:

In a Michelson interferometer, light of wavelength 632.8 nm from a He-Ne laser is used. When one of the mirrors is moved by a distance *D*, 8 fringes move past the field of view. What is the value of the distance *D*?

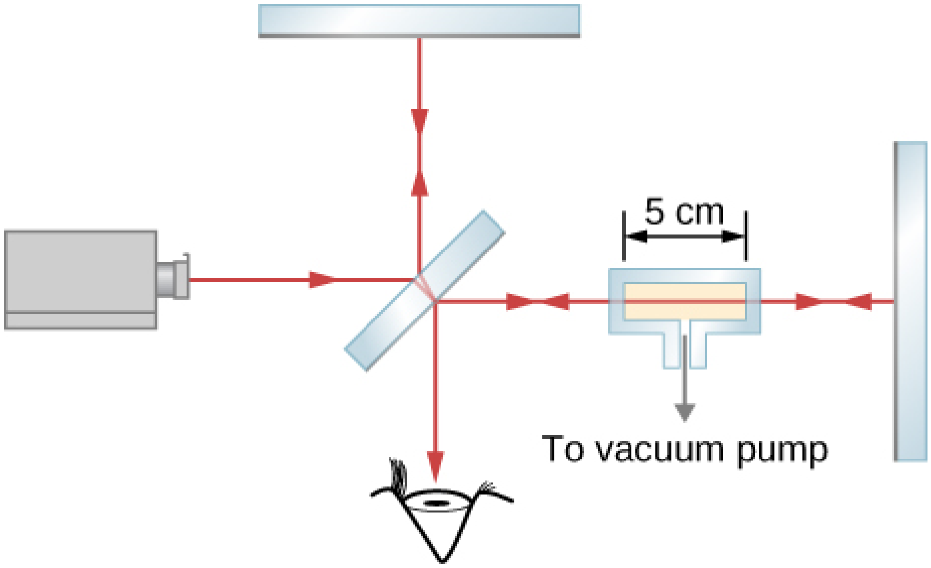
Solution:

$D = 2.53 \times 10^{-6}\,\text{m}$

Exercise:

Problem:

A chamber 5.0 cm long with flat, parallel windows at the ends is placed in one arm of a Michelson interferometer (see below). The light used has a wavelength of 500 nm in a vacuum. While all the air is being pumped out of the chamber, 29 fringes pass by a point on the observation screen. What is the refractive index of the air?



Additional Problems

Exercise:

Problem:

For 600-nm wavelength light and a slit separation of 0.12 mm, what are the angular positions of the first and third maxima in the double slit interference pattern?

Solution:

0.29° and 0.86°

Exercise:

Problem:

If the light source in the preceding problem is changed, the angular position of the third maximum is found to be 0.57°. What is the wavelength of light being used now?

Exercise:

Problem:

Red light ($\lambda = 710.\ \mathrm{nm}$) illuminates double slits separated by a distance $d = 0.150\ \mathrm{mm}$. The screen and the slits are 3.00 m apart. (a) Find the distance on the screen between the central maximum and the third maximum. (b) What is the distance between the second and the fourth maxima?

Solution:

a. 4.26 cm; b. 2.84 cm

Exercise:

Problem:

Two sources as in phase and emit waves with $\lambda = 0.42\ \mathrm{m}$. Determine whether constructive or destructive interference occurs at points whose distances from the two sources are (a) 0.84 and 0.42 m, (b) 0.21 and 0.42 m, (c) 1.26 and 0.42 m, (d) 1.87 and 1.45 m, (e) 0.63 and 0.84 m and (f) 1.47 and 1.26 m.

Exercise:

Problem:

Two slits $4.0 \times 10^{-6}\ \mathrm{m}$ apart are illuminated by light of wavelength 600 nm. What is the highest order fringe in the interference pattern?

Solution:

6

Exercise:

Problem:

Suppose that the highest order fringe that can be observed is the eighth in a double-slit experiment where 550-nm wavelength light is used. What is the minimum separation of the slits?

Exercise:

Problem:

The interference pattern of a He-Ne laser light $(\lambda = 632.9\,\text{nm})$ passing through two slits 0.031 mm apart is projected on a screen 10.0 m away. Determine the distance between the adjacent bright fringes.

Solution:

0.20 m

Exercise:

Problem:

Young’s double-slit experiment is performed immersed in water ($n = 1.333$). The light source is a He-Ne laser, $\lambda = 632.9\,\text{nm}$ in vacuum. (a) What is the wavelength of this light in water? (b) What is the angle for the third order maximum for two slits separated by 0.100 mm.

Exercise:

Problem:

A double-slit experiment is to be set up so that the bright fringes appear 1.27 cm apart on a screen 2.13 m away from the two slits. The light source was wavelength 500 nm. What should be the separation between the two slits?

Solution:

0.0839 mm

Exercise:

Problem:

An effect analogous to two-slit interference can occur with sound waves, instead of light. In an open field, two speakers placed 1.30 m apart are powered by a single-function generator producing sine waves at 1200-Hz frequency. A student walks along a line 12.5 m away and parallel to the line between the speakers. She hears an alternating pattern of loud and quiet, due to constructive and destructive interference. What is (a) the wavelength of this sound and (b) the distance between the central maximum and the first maximum (loud) position along this line?

Exercise:

Problem:

A hydrogen gas discharge lamp emits visible light at four wavelengths, $\lambda =$ 410, 434, 486, and 656 nm. (a) If light from this lamp falls on a N slits separated by 0.025 mm, how far from the central maximum are the third maxima when viewed on a screen 2.0 m from the slits? (b) By what distance are the second and third maxima separated for $l = 486\ \text{nm}$?

Solution:

a. 9.8, 10.4, 11.7, and 15.7 cm; b. 3.9 cm

Exercise:

Problem:

Monochromatic light of frequency $5.5\ \times\ 10^{14}\ \text{Hz}$ falls on 10 slits separated by 0.020 mm. What is the separation between the first and third maxima on a screen that is 2.0 m from the slits?

Exercise:

Problem:

Eight slits equally separated by 0.149 mm is uniformly illuminated by a monochromatic light at $\lambda = 523\ \text{nm}$. What is the width of the central principal maximum on a screen 2.35 m away?

Solution:

0.0575°

Exercise:

Problem:

Eight slits equally separated by 0.149 mm is uniformly illuminated by a monochromatic light at $\lambda = 523\ \text{nm}$. What is the intensity of a secondary maxima compared to that of the principal maxima?

Exercise:

Problem:

A transparent film of thickness 250 nm and index of refraction of 1.40 is surrounded by air. What wavelength in a beam of white light at near-normal incidence to the film undergoes destructive interference when reflected?

Solution:

700 nm

Exercise:

Problem:

An intensity minimum is found for 450 nm light transmitted through a transparent film $(n = 1.20)$ in air. (a) What is minimum thickness of the film? (b) If this wavelength is the longest for which the intensity minimum occurs, what are the next three lower values of λ for which this happens?

Exercise:

Problem:

A thin film with $n = 1.32$ is surrounded by air. What is the minimum thickness of this film such that the reflection of normally incident light with $\lambda = 500\ \mathrm{nm}$ is minimized?

Solution:

189 nm

Exercise:

Problem:

Repeat your calculation of the previous problem with the thin film placed on a flat glass ($n = 1.50$) surface.

Exercise:

Problem:

After a minor oil spill, a think film of oil ($n = 1.40$) of thickness 450 nm floats on the water surface in a bay. (a) What predominant color is seen by a bird flying overhead? (b) What predominant color is seen by a seal swimming underwater?

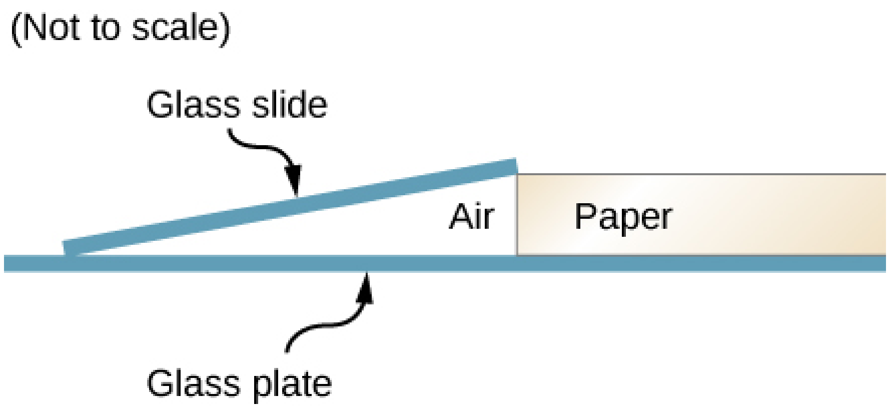
Solution:

a. green (504 nm); b. magenta (white minus green)

Exercise:

Problem:

A microscope slide 10 cm long is separated from a glass plate at one end by a sheet of paper. As shown below, the other end of the slide is in contact with the plate. The slide is illuminated from above by light from a sodium lamp ($\lambda = 589\ \mathrm{nm}$), and 14 fringes per centimeter are seen along the slide. What is the thickness of the piece of paper?



Exercise:

Problem:

Suppose that the setup of the preceding problem is immersed in an unknown liquid. If 18 fringes per centimeter are now seen along the slide, what is the index of refraction of the liquid?

Solution:

1.29

Exercise:

Problem:

A thin wedge filled with air is produced when two flat glass plates are placed on top of one another and a slip of paper is inserted between them at one edge. Interference fringes are observed when monochromatic light falling vertically on the plates are seen in reflection. Is the first fringe near the edge where the plates are in contact a bright fringe or a dark fringe? Explain.

Exercise:

Problem:

Two identical pieces of rectangular plate glass are used to measure the thickness of a hair. The glass plates are in direct contact at one edge and a single hair is placed between them hear the opposite edge. When illuminated with a sodium lamp ($\lambda = 589\ \text{nm}$), the hair is seen between the 180th and 181st dark fringes. What are the lower and upper limits on the hair's diameter?

Solution:

$52.7\ \mu\text{m}$ and $53.0\ \mu\text{m}$

Exercise:

Problem:

Two microscope slides made of glass are illuminated by monochromatic ($\lambda = 589\ \text{nm}$) light incident perpendicularly. The top slide touches the bottom slide at one end and rests on a thin copper wire at the other end, forming a wedge of air. The diameter of the copper wire is $29.45\ \mu\text{m}$. How many bright fringes are seen across these slides?

Exercise:

Problem:

A good quality camera "lens" is actually a system of lenses, rather than a single lens, but a side effect is that a reflection from the surface of one lens can bounce around many times within the system, creating artifacts in the photograph. To counteract this problem, one of the lenses in such a system is coated with a thin layer of material ($n = 1.28$) on one side. The index of refraction of the lens glass is 1.68. What is the smallest thickness of the coating that reduces the reflection at 640 nm by destructive interference? (In other words, the coating's effect is to be optimized for $\lambda = 640\ \text{nm}$.)

Solution:

160 nm

Exercise:

Problem:

Constructive interference is observed from directly above an oil slick for wavelengths (in air) 440 nm and 616 nm. The index of refraction of this oil is $n = 1.54$. What is the film's minimum possible thickness?

Exercise:

Problem:

A soap bubble is blown outdoors. What colors (indicate by wavelengths) of the reflected sunlight are seen enhanced? The soap bubble has index of refraction 1.36 and thickness 380 nm.

Solution:

413 nm and 689 nm

Exercise:

Problem:

A Michelson interferometer with a He-Ne laser light source ($\lambda = 632.8\ \text{nm}$) projects its interference pattern on a screen. If the movable mirror is caused to move by $8.54\ \mu\text{m}$, how many fringes will be observed shifting through a reference point on a screen?

Exercise:

Problem:

An experimenter detects 251 fringes when the movable mirror in a Michelson interferometer is displaced. The light source used is a sodium lamp, wavelength 589 nm. By what distance did the movable mirror move?

Solution:

$73.9\ \mu\text{m}$

Exercise:

Problem:

A Michelson interferometer is used to measure the wavelength of light put through it. When the movable mirror is moved by exactly 0.100 mm, the number of fringes observed moving through is 316. What is the wavelength of the light?

Exercise:

Problem:

A 5.08-cm-long rectangular glass chamber is inserted into one arm of a Michelson interferometer using a 633-nm light source. This chamber is initially filled with air $(n = 1.000293)$ at standard atmospheric pressure but the air is gradually pumped out using a vacuum pump until a near perfect vacuum is achieved. How many fringes are observed moving by during the transition?

Solution:

47

Exercise:

Problem:

Into one arm of a Michelson interferometer, a plastic sheet of thickness $75\ \mu\text{m}$ is inserted, which causes a shift in the interference pattern by 86 fringes. The light source has wavelength of 610 nm in air. What is the index of refraction of this plastic?

Exercise:

Problem:

The thickness of an aluminum foil is measured using a Michelson interferometer that has its movable mirror mounted on a micrometer. There is a difference of 27 fringes in the observed interference pattern when the micrometer clamps down on the foil compared to when the micrometer is empty. Calculate the thickness of the foil?

Solution:

$8.5\ \mu\text{m}$

Exercise:

Problem:

The movable mirror of a Michelson interferometer is attached to one end of a thin metal rod of length 23.3 mm. The other end of the rod is anchored so it does not move. As the temperature of the rod changes from $15\ ^{\circ}\text{C}$ to $25\ \text{C}$, a change of 14 fringes is observed. The light source is a He Ne laser, $\lambda = 632.8\ \text{nm}$. What is the change in length of the metal bar, and what is its thermal expansion coefficient?

Exercise:

Problem:

In a thermally stabilized lab, a Michelson interferometer is used to monitor the temperature to ensure it stays constant. The movable mirror is mounted on the end of a 1.00-m-long aluminum rod, held fixed at the other end. The light source is a He Ne laser, $\lambda = 632.8\ \text{nm}$. The resolution of this apparatus corresponds to the temperature difference when a change of just one fringe is observed. What is this temperature difference?

Solution:

0.013°C

Exercise:

Problem:

A 65-fringe shift results in a Michelson interferometer when a $42.0\text{-}\mu\text{m}$ film made of an unknown material is placed in one arm. The light source has wavelength 632.9 nm. Identify the material using the indices of refraction found in [link].

Challenge Problems

Exercise:

Problem:

Determine what happens to the double-slit interference pattern if one of the slits is covered with a thin, transparent film whose thickness is $\lambda/[2(n-1)]$, where λ is the wavelength of the incident light and *n* is the index of refraction of the film.

Solution:

Bright and dark fringes switch places.

Exercise:

Problem:

Fifty-one narrow slits are equally spaced and separated by 0.10 mm. The slits are illuminated by blue light of wavelength 400 nm. What is angular position of the twenty-fifth secondary maximum? What is its peak intensity in comparison with that of the primary maximum?

Exercise:

Problem:

A film of oil on water will appear dark when it is very thin, because the path length difference becomes small compared with the wavelength of light and there is a phase shift at the top surface. If it becomes dark when the path length difference is less than one-fourth the wavelength, what is the thickest the oil can be and appear dark at all visible wavelengths? Oil has an index of refraction of 1.40.

Solution:

The path length must be less than one-fourth of the shortest visible wavelength in oil. The thickness of the oil is half the path length, so it must be less than one-eighth of the shortest visible wavelength in oil. If we take 380 nm to be the shortest visible wavelength in air, 33.9 nm.

Exercise:

Problem:

[link] shows two glass slides illuminated by monochromatic light incident perpendicularly. The top slide touches the bottom slide at one end and rests on a 0.100-mm-diameter hair at the other end, forming a wedge of air. (a) How far apart are the dark bands, if the slides are 7.50 cm long and 589-nm light is used? (b) Is there any difference if the slides are made from crown or flint glass? Explain.

Exercise:

Problem:

[link] shows two 7.50-cm-long glass slides illuminated by pure 589-nm wavelength light incident perpendicularly. The top slide touches the bottom slide at one end and rests on some debris at the other end, forming a wedge of air. How thick is the debris, if the dark bands are 1.00 mm apart?

Solution:

$4.42 \times 10^{-5}\,\text{m}$

Exercise:

Problem:

A soap bubble is 100 nm thick and illuminated by white light incident at a 45° angle to its surface. What wavelength and color of visible light is most constructively reflected, assuming the same index of refraction as water?

Exercise:

Problem:

An oil slick on water is 120 nm thick and illuminated by white light incident at a 45° angle to its surface. What color does the oil appear (what is the most constructively reflected wavelength), given its index of refraction is 1.40?

Solution:

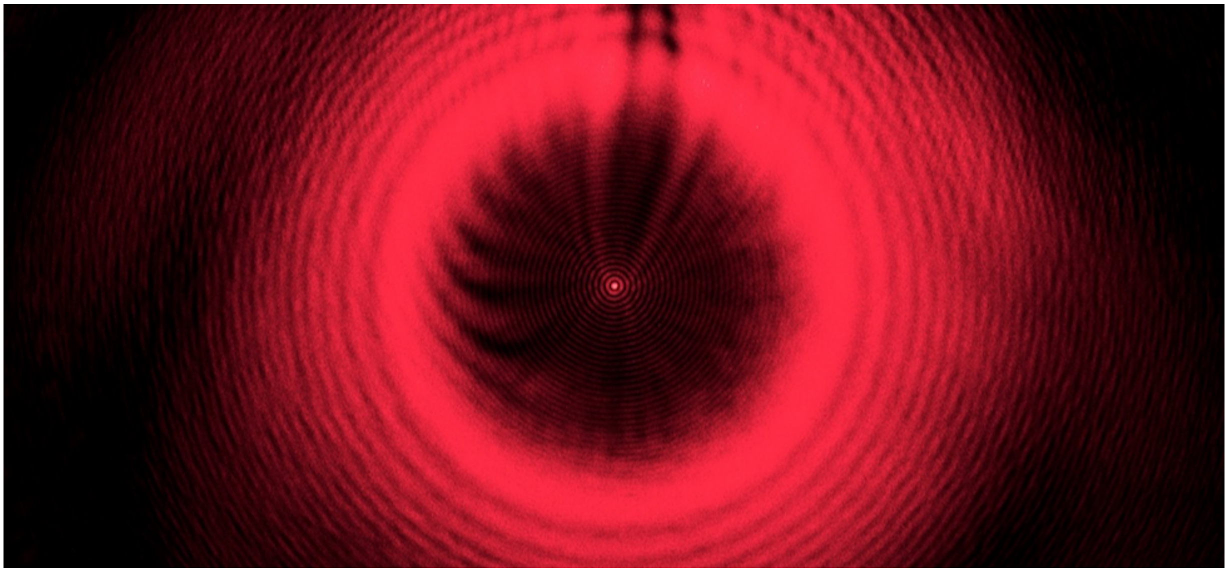
for one phase change: 950 nm (infrared); for three phase changes: 317 nm (ultraviolet); Therefore, the oil film will appear black, since the reflected light is not in the visible part of the spectrum.

Glossary

interferometer	instrument that uses interference of waves to make measurements

Introduction class="introduction"

A steel ball bearing illuminated by a laser does not cast a sharp, circular shadow. Instead, a series of diffraction fringes and a central bright spot are observed. Known as Poisson’s spot, the effect was first predicted by Augustin-Jean Fresnel (1788–1827) as a consequence of diffraction of light waves. Based on principles of ray optics, Siméon-Denis Poisson (1781–1840) argued against Fresnel’s prediction. (credit: modification of work by Harvard Natural



Imagine passing a monochromatic light beam through a narrow opening—a slit just a little wider than the wavelength of the light. Instead of a simple shadow of the slit on the screen, you will see that an interference pattern appears, even though there is only one slit.

In the chapter on interference, we saw that you need two sources of waves for interference to occur. How can there be an interference pattern when we have only one slit? In The Nature of Light, we learned that, due to Huygens's principle, we can imagine a wave front as equivalent to infinitely many point sources of waves. Thus, a wave from a slit can behave not as one wave but as an infinite number of point sources. These waves can interfere with each other, resulting in an interference pattern without the presence of a second slit. This phenomenon is called *diffraction*.

Another way to view this is to recognize that a slit has a small but finite width. In the preceding chapter, we implicitly regarded slits as objects with positions but no size. The widths of the slits were considered negligible. When the slits have finite widths, each point along the opening can be considered a point source of light—a foundation of Huygens's principle.

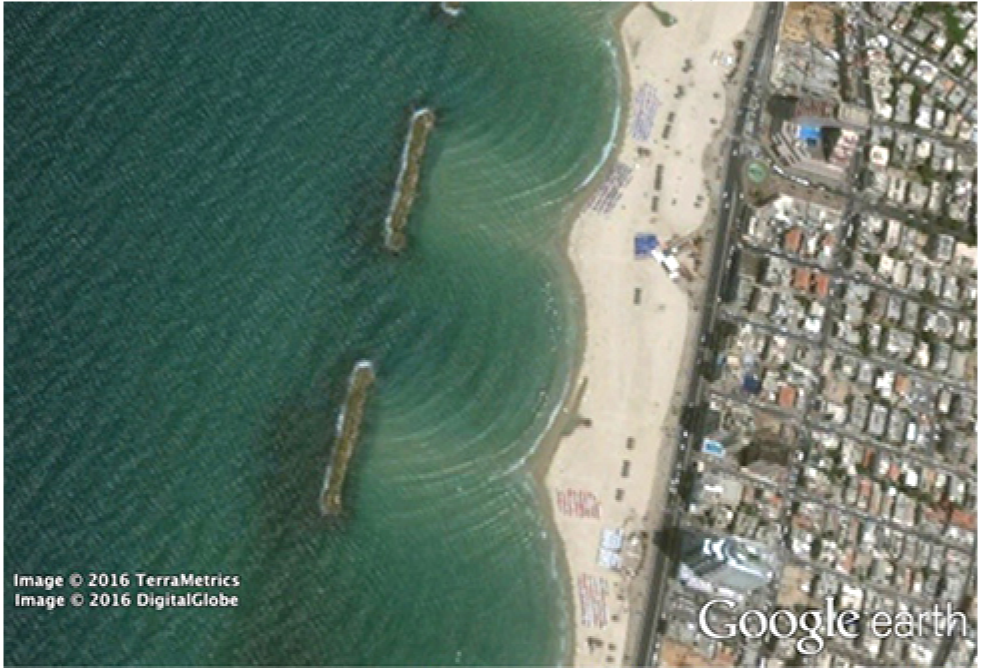
Because real-world optical instruments must have finite apertures (otherwise, no light can enter), diffraction plays a major role in the way we interpret the output of these optical instruments. For example, diffraction places limits on our ability to resolve images or objects. This is a problem that we will study later in this chapter.

Single-Slit Diffraction

By the end of this section, you will be able to:

- Explain the phenomenon of diffraction and the conditions under which it is observed
- Describe diffraction through a single slit

After passing through a narrow aperture (opening), a wave propagating in a specific direction tends to spread out. For example, sound waves that enter a room through an open door can be heard even if the listener is in a part of the room where the geometry of ray propagation dictates that there should only be silence. Similarly, ocean waves passing through an opening in a breakwater can spread throughout the bay inside. ([link]). The spreading and bending of sound and ocean waves are two examples of **diffraction**, which is the bending of a wave around the edges of an opening or an obstacle—a phenomenon exhibited by all types of waves.

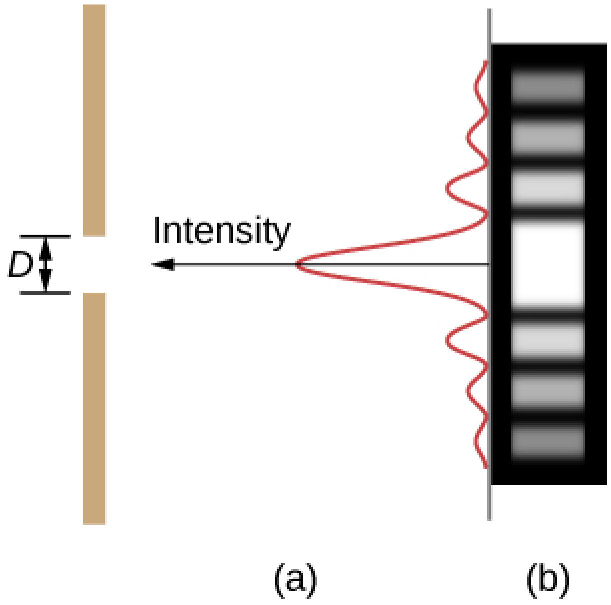


Because of the diffraction of waves, ocean waves entering through an opening in a breakwater can spread throughout the bay. (credit: modification of map data from Google Earth)

The diffraction of sound waves is apparent to us because wavelengths in the audible region are approximately the same size as the objects they encounter, a condition that must be satisfied if diffraction effects are to be observed easily. Since the wavelengths of visible light range from approximately 390 to 770 nm, most objects do not diffract light significantly. However, situations do occur in which apertures are small enough that the diffraction of light is observable. For example, if you place your middle and index fingers close together and look through the opening at a light bulb, you can see a rather clear diffraction pattern, consisting of light and dark lines running parallel to your fingers.

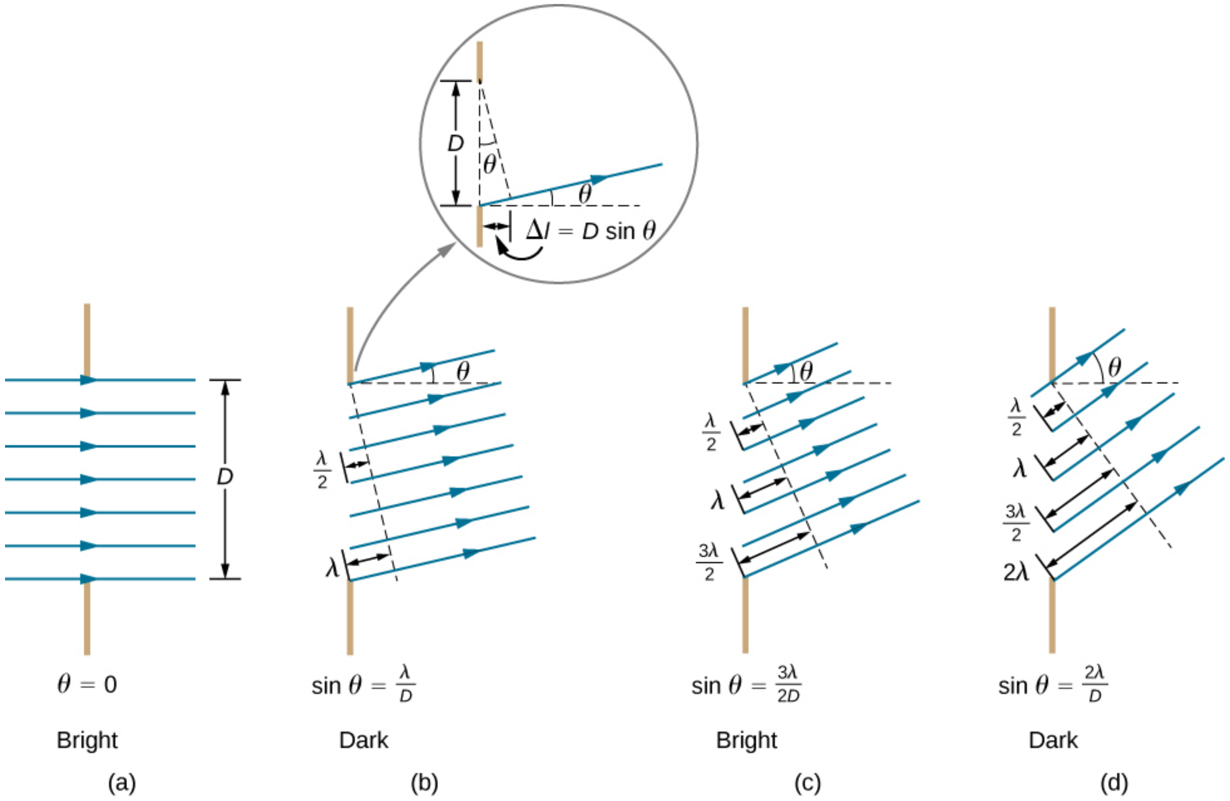
Diffraction through a Single Slit

Light passing through a single slit forms a diffraction pattern somewhat different from those formed by double slits or diffraction gratings, which we discussed in the chapter on interference. [link] shows a single-slit diffraction pattern. Note that the central maximum is larger than maxima on either side and that the intensity decreases rapidly on either side. In contrast, a diffraction grating (Diffraction Gratings) produces evenly spaced lines that dim slowly on either side of the center.



Single-slit diffraction pattern. (a) Monochromatic light passing through a single slit has a central maximum and many smaller and dimmer maxima on either side. The central maximum is six times higher than shown. (b) The diagram shows the bright central maximum, and the dimmer and thinner maxima on either side.

The analysis of single-slit diffraction is illustrated in [link]. Here, the light arrives at the slit, illuminating it uniformly and is in phase across its width. We then consider light propagating onwards from different parts of the *same* slit. According to Huygens's principle, every part of the wave front in the slit emits wavelets, as we discussed in The Nature of Light. These are like rays that start out in phase and head in all directions. (Each ray is perpendicular to the wave front of a wavelet.) Assuming the screen is very far away compared with the size of the slit, rays heading toward a common destination are nearly parallel. When they travel straight ahead, as in part (a) of the figure, they remain in phase, and we observe a central maximum. However, when rays travel at an angle θ relative to the original direction of the beam, each ray travels a different distance to a common location, and they can arrive in or out of phase. In part (b), the ray from the bottom travels a distance of one wavelength λ farther than the ray from the top. Thus, a ray from the center travels a distance $\lambda/2$ less than the one at the bottom edge of the slit, arrives out of phase, and interferes destructively. A ray from slightly above the center and one from slightly above the bottom also cancel one another. In fact, each ray from the slit interferes destructively with another ray. In other words, a pair-wise cancellation of all rays results in a dark minimum in intensity at this angle. By symmetry, another minimum occurs at the same angle to the right of the incident direction (toward the bottom of the figure) of the light.



Light passing through a single slit is diffracted in all directions and may interfere constructively or destructively, depending on the angle. The difference in path length for rays from either side of the slit is seen to be $D \sin \theta$.

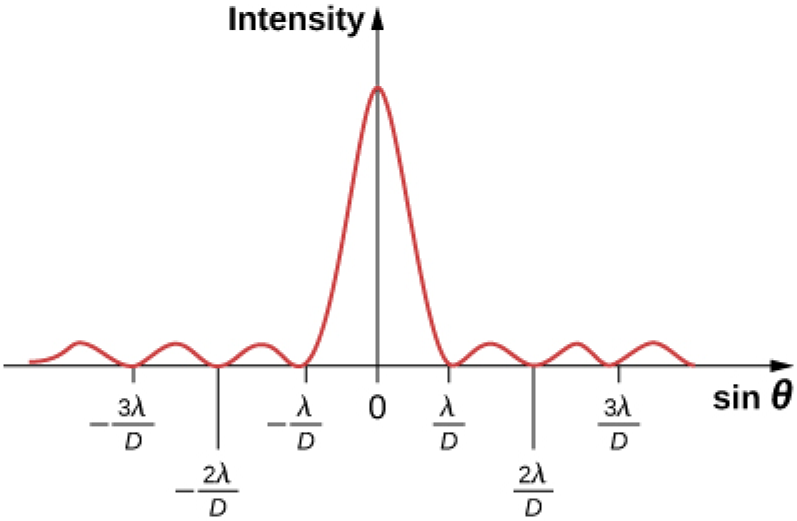
At the larger angle shown in part (c), the path lengths differ by $3\lambda/2$ for rays from the top and bottom of the slit. One ray travels a distance λ different from the ray from the bottom and arrives in phase, interfering constructively. Two rays, each from slightly above those two, also add constructively. Most rays from the slit have another ray to interfere with constructively, and a maximum in intensity occurs at this angle. However, not all rays interfere constructively for this situation, so the maximum is not as intense as the central maximum. Finally, in part (d), the angle shown is large enough to produce a second minimum. As seen in the figure, the difference in path length for rays from either side of the slit is $D \sin \theta$, and we see that a destructive minimum is obtained when this distance is an integral multiple of the wavelength.

Thus, to obtain **destructive interference for a single slit**,

Note:
Equation:

$$D\sin\theta = m\lambda,\ \text{for}\ m = \pm 1, \pm 2, \pm 3, \ldots \text{(destructive)},$$

where D is the slit width, λ is the light's wavelength, θ is the angle relative to the original direction of the light, and m is the order of the minimum. [link] shows a graph of intensity for single-slit interference, and it is apparent that the maxima on either side of the central maximum are much less intense and not as wide. This effect is explored in Double-Slit Diffraction.

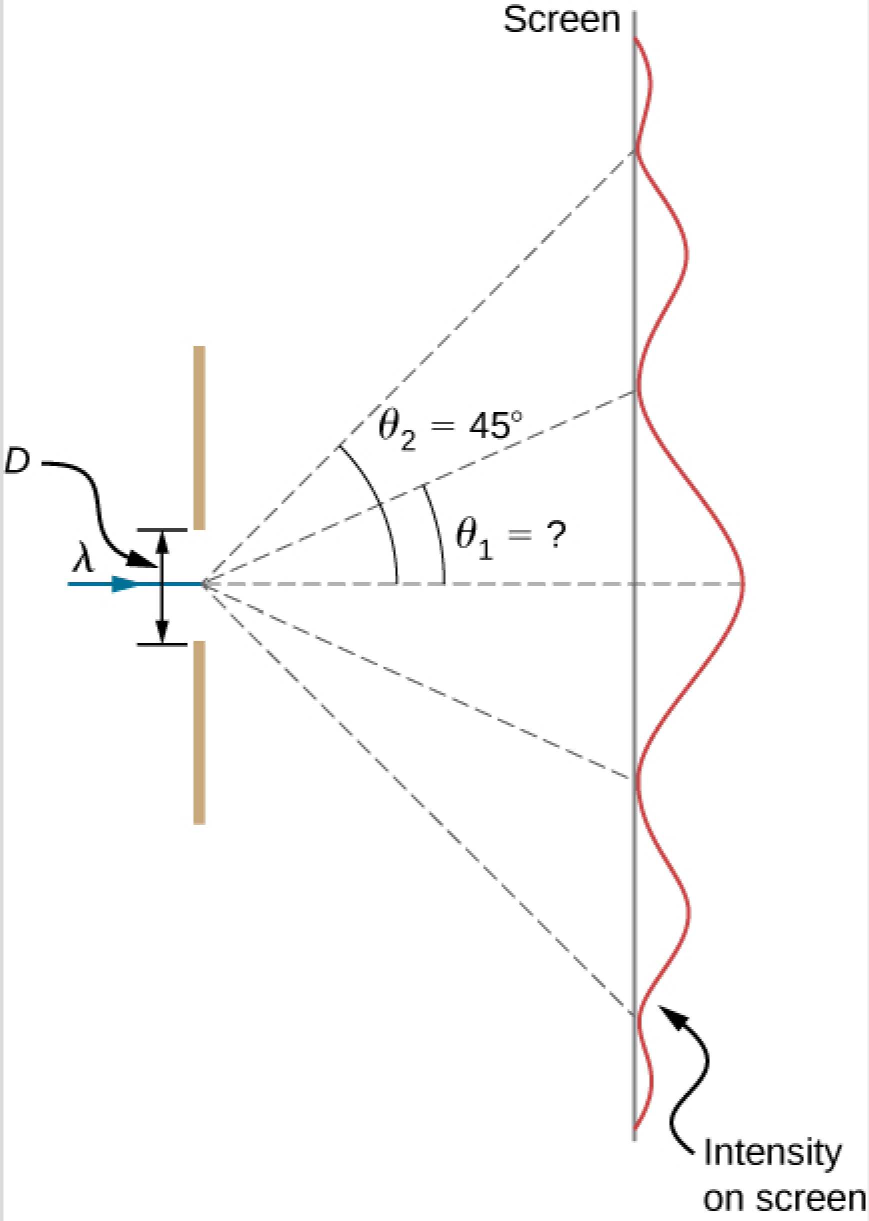


A graph of single-slit diffraction intensity showing the central maximum to be wider and much more intense than those to the sides. In fact, the central maximum is six times higher than shown here.

Example:

Calculating Single-Slit Diffraction

Visible light of wavelength 550 nm falls on a single slit and produces its second diffraction minimum at an angle of 45.0° relative to the incident direction of the light, as in [link]. (a) What is the width of the slit? (b) At what angle is the first minimum produced?



In this example, we analyze a graph of the single-slit diffraction pattern.

Strategy

From the given information, and assuming the screen is far away from the slit, we can use the equation $D\sin\theta = m\lambda$ first to find D, and again to find the angle for the first minimum θ_1.

Solution

a. We are given that λ $= 550\,\text{nm}$, $m = 2$, and $\theta_2 = 45.0^\circ$. Solving the equation $D\sin\theta = m\lambda$ for D and substituting known values gives
   Equation:

$$D = \frac{m\lambda}{\sin\theta_2} = \frac{2(550\,\text{nm})}{\sin 45.0^\circ} = \frac{1100\ \times\ 10^{-9}\,\text{m}}{0.707} = 1.56\ \times\ 10^{-6}\,\text{m}.$$

b. Solving the equation $D\sin\theta = m\lambda$ for $\sin\theta_1$ and substituting the known values gives
   Equation:

$$\sin\theta_1 = \frac{m\lambda}{D} = \frac{1(550\ \times\ 10^{-9}\,\text{m})}{1.56\ \times\ 10^{-6}\,\text{m}}.$$

   Thus the angle θ_1 is
   Equation:

$$\theta_1 = \sin^{-1}0.354 = 20.7^\circ.$$

Significance

We see that the slit is narrow (it is only a few times greater than the wavelength of light). This is consistent with the fact that light must interact with an object comparable in size to its wavelength in order to exhibit significant wave effects such as this single-slit diffraction pattern. We also see that the central maximum extends 20.7° on either side of the original beam, for a width of about 41°. The angle between the first and second minima is only about $24^\circ (45.0^\circ - 20.7^\circ)$. Thus, the second maximum is only about half as wide as the central maximum.

Note:
Exercise:

Problem:

Check Your Understanding Suppose the slit width in [link] is increased to $1.8 \times 10^{-6}\,\text{m}$. What are the new angular positions for the first, second, and third minima? Would a fourth minimum exist?

Solution:

17.8°, 37.7°, 66.4°; no

Summary

- Diffraction can send a wave around the edges of an opening or other obstacle.
- A single slit produces an interference pattern characterized by a broad central maximum with narrower and dimmer maxima to the sides.

Conceptual Questions

Exercise:

Problem:

As the width of the slit producing a single-slit diffraction pattern is reduced, how will the diffraction pattern produced change?

Solution:

The diffraction pattern becomes wider.

Exercise:

Problem: Compare interference and diffraction.

Exercise:

Problem:

If you and a friend are on opposite sides of a hill, you can communicate with walkie-talkies but not with flashlights. Explain.

Solution:

Walkie-talkies use radio waves whose wavelengths are comparable to the size of the hill and are thus able to diffract around the hill. Visible wavelengths of the flashlight travel as rays at this size scale.

Exercise:

Problem:

What happens to the diffraction pattern of a single slit when the entire optical apparatus is immersed in water?

Exercise:

Problem:

In our study of diffraction by a single slit, we assume that the length of the slit is much larger than the width. What happens to the diffraction pattern if these two dimensions were comparable?

Solution:

The diffraction pattern becomes two-dimensional, with main fringes, which are now spots, running in perpendicular directions and fainter spots in intermediate directions.

Exercise:

Problem:

A rectangular slit is twice as wide as it is high. Is the central diffraction peak wider in the vertical direction or in the horizontal direction?

Problems

Exercise:

Problem:

(a) At what angle is the first minimum for 550-nm light falling on a single slit of width $1.00\mu\mathrm{m}$? (b) Will there be a second minimum?

Solution:

a. 33.4°; b. no

Exercise:

Problem:

(a) Calculate the angle at which a 2.00-$\mu\mathrm{m}$-wide slit produces its first minimum for 410-nm violet light. (b) Where is the first minimum for 700-nm red light?

Exercise:

Problem:

(a) How wide is a single slit that produces its first minimum for 633-nm light at an angle of 28.0° ? (b) At what angle will the second minimum be?

Solution:

a. $1.35\ \times\ 10^{-6}\,\mathrm{m}$; b. 69.9°

Exercise:

Problem:

(a) What is the width of a single slit that produces its first minimum at 60.0° for 600-nm light? (b) Find the wavelength of light that has its first minimum at 62.0°.

Exercise:

Problem:

Find the wavelength of light that has its third minimum at an angle of 48.6° when it falls on a single slit of width $3.00 \mu\text{m}$.

Solution:

750 nm

Exercise:

Problem:

(a) Sodium vapor light averaging 589 nm in wavelength falls on a single slit of width $7.50 \mu\text{m}$. At what angle does it produces its second minimum? (b) What is the highest-order minimum produced?

Exercise:

Problem:

Consider a single-slit diffraction pattern for $\lambda = 589 \text{ nm}$, projected on a screen that is 1.00 m from a slit of width 0.25 mm. How far from the center of the pattern are the centers of the first and second dark fringes?

Solution:

2.4 mm, 4.7 mm

Exercise:

Problem:

(a) Find the angle between the first minima for the two sodium vapor lines, which have wavelengths of 589.1 and 589.6 nm, when they fall upon a single slit of width $2.00 \mu m$. (b) What is the distance between these minima if the diffraction pattern falls on a screen 1.00 m from the slit? (c) Discuss the ease or difficulty of measuring such a distance.

Exercise:

Problem:

(a) What is the minimum width of a single slit (in multiples of λ) that will produce a first minimum for a wavelength λ ? (b) What is its minimum width if it produces 50 minima? (c) 1000 minima?

Solution:

a. 1.00λ; b. 50.0λ; c. 1000λ

Exercise:

Problem:

(a) If a single slit produces a first minimum at 14.5°, at what angle is the second-order minimum? (b) What is the angle of the third-order minimum? (c) Is there a fourth-order minimum? (d) Use your answers to illustrate how the angular width of the central maximum is about twice the angular width of the next maximum (which is the angle between the first and second minima).

Exercise:

Problem:

If the separation between the first and the second minima of a single-slit diffraction pattern is 6.0 mm, what is the distance between the screen and the slit? The light wavelength is 500 nm and the slit width is 0.16 mm.

Solution:

1.92 m

Exercise:

Problem:

A water break at the entrance to a harbor consists of a rock barrier with a 50.0-m-wide opening. Ocean waves of 20.0-m wavelength approach the opening straight on. At what angles to the incident direction are the boats inside the harbor most protected against wave action?

Exercise:

Problem:

An aircraft maintenance technician walks past a tall hangar door that acts like a single slit for sound entering the hangar. Outside the door, on a line perpendicular to the opening in the door, a jet engine makes a 600-Hz sound. At what angle with the door will the technician observe the first minimum in sound intensity if the vertical opening is 0.800 m wide and the speed of sound is 340 m/s?

Solution:

45.1°

Glossary

destructive interference for a single slit
occurs when the width of the slit is comparable to the wavelength of light illuminating it

diffraction
bending of a wave around the edges of an opening or an obstacle

Intensity in Single-Slit Diffraction

By the end of this section, you will be able to:

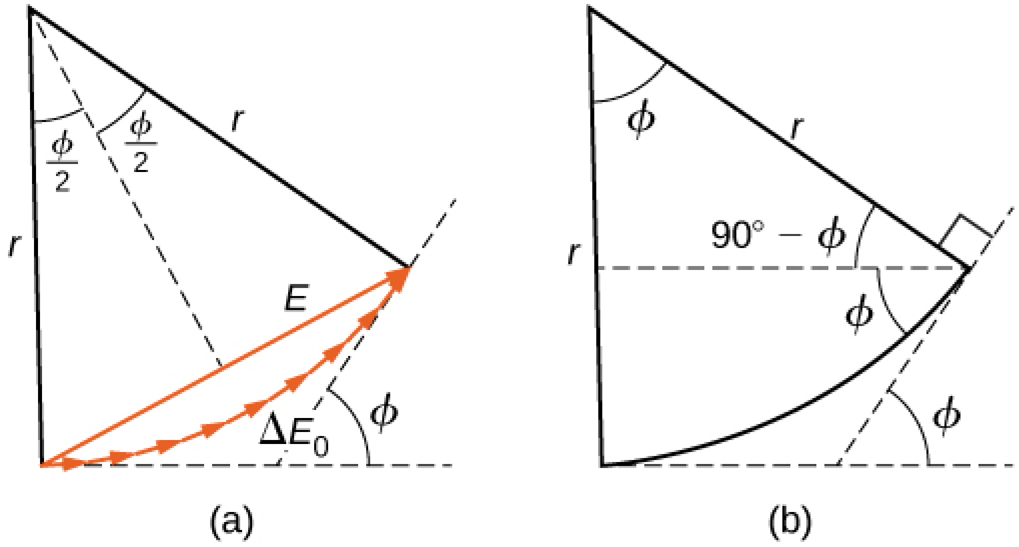
- Calculate the intensity relative to the central maximum of the single-slit diffraction peaks
- Calculate the intensity relative to the central maximum of an arbitrary point on the screen

To calculate the intensity of the diffraction pattern, we follow the phasor method used for calculations with ac circuits in Alternating-Current Circuits. If we consider that there are N Huygens sources across the slit shown in [link], with each source separated by a distance D/N from its adjacent neighbors, the path difference between waves from adjacent sources reaching the arbitrary point P on the screen is $(D/N)\sin\theta$. This distance is equivalent to a phase difference of $(2\pi D/\lambda N)\sin\theta$. The phasor diagram for the waves arriving at the point whose angular position is θ is shown in [link]. The amplitude of the phasor for each Huygens wavelet is ΔE_0, the amplitude of the resultant phasor is E, and the phase difference between the wavelets from the first and the last sources is

Equation:

$$\phi = \left(\frac{2\pi}{\lambda}\right) D \sin\theta.$$

With $N \to \infty$, the phasor diagram approaches a circular arc of length $N\Delta E_0$ and radius r. Since the length of the arc is $N\Delta E_0$ for any ϕ, the radius r of the arc must decrease as ϕ increases (or equivalently, as the phasors form tighter spirals).



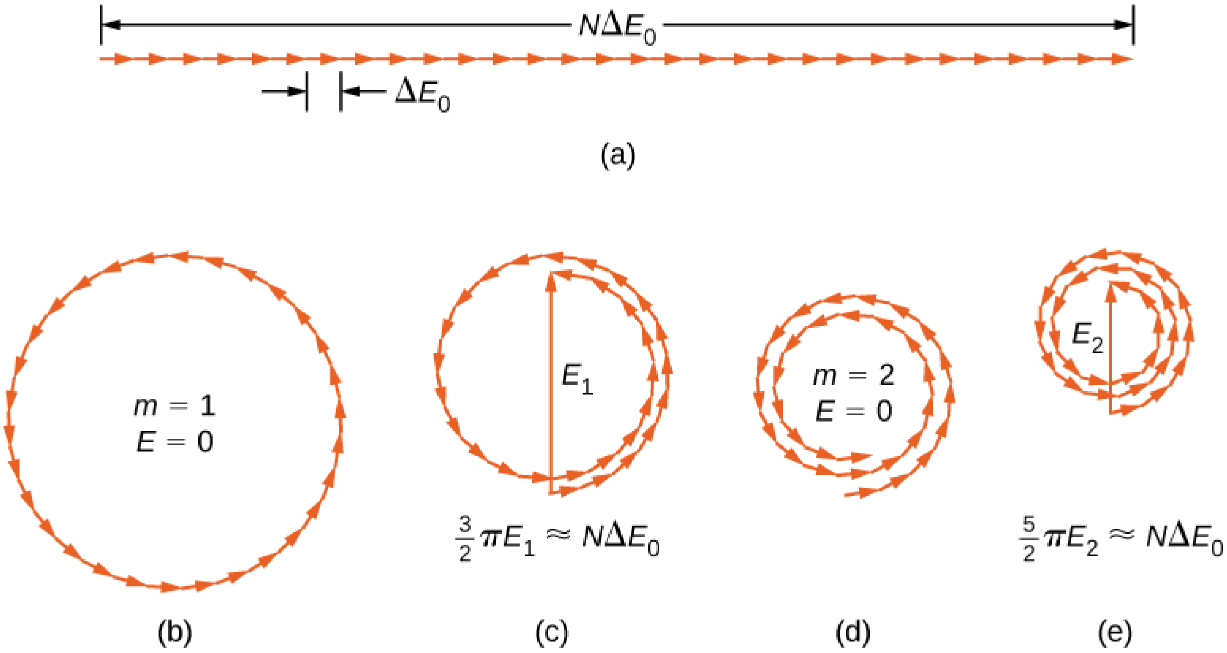
(a) Phasor diagram corresponding to the angular position θ in the single-slit diffraction pattern. The phase difference between the wavelets from the first and last sources is $\phi = (2\pi/\lambda)D\sin\theta$. (b) The geometry of the phasor diagram.

The phasor diagram for $\phi = 0$ (the center of the diffraction pattern) is shown in [link](a) using $N = 30$. In this case, the phasors are laid end to end in a straight line of length $N\Delta E_0$, the radius *r* goes to infinity, and the resultant has its maximum value $E = N\Delta E_0$. The intensity of the light can be obtained using the relation $I = \frac{1}{2}c\varepsilon_0 E^2$ from Electromagnetic Waves. The intensity of the maximum is then

Equation:

$$I_0 = \frac{1}{2}c\varepsilon_0(N\Delta E_0)^2 = \frac{1}{2\mu_0 c}(N\Delta E_0)^2,$$

where $\varepsilon_0 = 1/\mu_0 c^2$. The phasor diagrams for the first two zeros of the diffraction pattern are shown in parts (b) and (d) of the figure. In both cases, the phasors add to zero, after rotating through $\phi = 2\pi$ rad for $m = 1$ and 4π rad for $m = 2$.


$N\Delta E_0$
ΔE_0
(a)
$m = 1$
$E = 0$
E_1
$\frac{3}{2}\pi E_1 \approx N\Delta E_0$
$m = 2$
$E = 0$
E_2
$\frac{5}{2}\pi E_2 \approx N\Delta E_0$
(b)
(c)
(d)
(e)

Phasor diagrams (with 30 phasors) for various points on the single-slit diffraction pattern. Multiple rotations around a given circle have been separated slightly so that the phasors can be seen. (a) Central maximum, (b) first minimum, (c) first maximum beyond central maximum, (d) second minimum, and (e) second maximum beyond central maximum.

The next two maxima beyond the central maxima are represented by the phasor diagrams of parts (c) and (e). In part (c), the phasors have rotated through $\phi = 3\pi$ rad and have formed a resultant phasor of magnitude E_1. The length of the arc formed by the phasors is $N\Delta E_0$. Since this corresponds to 1.5 rotations around a circle of diameter E_1, we have

Equation:

$$\frac{3}{2}\pi E_1 \approx N\Delta E_0,$$

so

Equation:

$$E_1 = \frac{2N\Delta E_0}{3\pi}$$

and
Equation:

$$I_1 = \frac{1}{2\mu_0 c} E_1^2 = \frac{4(N\Delta E_0)^2}{(9\pi^2)\,(2\mu_0 c)} \approx 0.045 I_0,$$

where
Equation:

$$I_0 = \frac{(N\Delta E_0)^2}{2\mu_0 c}.$$

In part (e), the phasors have rotated through $\phi = 5\pi$ rad, corresponding to 2.5 rotations around a circle of diameter E_2 and arc length $N\Delta E_0$. This results in $I_2 \approx 0.016 I_0$. The proof is left as an exercise for the student ([link]).

These two maxima actually correspond to values of ϕ slightly less than 3π rad and 5π rad. Since the total length of the arc of the phasor diagram is always $N\Delta E_0$, the radius of the arc decreases as ϕ increases. As a result, E_1 and E_2 turn out to be slightly larger for arcs that have not quite curled through 3π rad and 5π rad, respectively. The exact values of ϕ for the maxima are investigated in [link]. In solving that problem, you will find that they are less than, but very close to, $\phi = 3\pi, 5\pi, 7\pi, \ldots$ rad.

To calculate the intensity at an arbitrary point *P* on the screen, we return to the phasor diagram of [link]. Since the arc subtends an angle ϕ at the center of the circle,
Equation:

$$N\Delta E_0 = r\phi$$

and
Equation:

$$\sin\left(\frac{\phi}{2}\right)=\frac{E}{2r}.$$

where E is the amplitude of the resultant field. Solving the second equation for E and then substituting r from the first equation, we find
Equation:

$$E=2r\sin\frac{\phi}{2}=2\frac{N\Delta E_o}{\phi}\sin\frac{\phi}{2}.$$

Now defining

Note:
Equation:

$$\beta=\frac{\phi}{2}=\frac{\pi D\sin\theta}{\lambda}$$

we obtain

Note:
Equation:

$$E=N\Delta E_0\frac{\sin\beta}{\beta}$$

This equation relates the amplitude of the resultant field at any point in the diffraction pattern to the amplitude $N\Delta E_0$ at the central maximum. The intensity is proportional to the square of the amplitude, so

Note:
Equation:

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2$$

where $I_0 = (N\Delta E_0)^2 / 2\mu_0 c$ is the intensity at the center of the pattern.

For the central maximum, $\phi = 0$, β is also zero and we see from l'Hôpital's rule that $\lim_{\beta \to 0} (\sin \beta / \beta) = 1$, so that $\lim_{\phi \to 0} I = I_0$. For the next maximum, $\phi = 3\pi$ rad, we have $\beta = 3\pi/2$ rad and when substituted into [link], it yields

Equation:

$$I_1 = I_0 \left(\frac{\sin 3\pi/2}{3\pi/2} \right)^2 \approx 0.045 I_0,$$

in agreement with what we found earlier in this section using the diameters and circumferences of phasor diagrams. Substituting $\phi = 5\pi$ rad into [link] yields a similar result for I_2.

A plot of [link] is shown in [link] and directly below it is a photograph of an actual diffraction pattern. Notice that the central peak is much brighter than the others, and that the zeros of the pattern are located at those points where $\sin \beta = 0$, which occurs when $\beta = m\pi$ rad. This corresponds to

Equation:

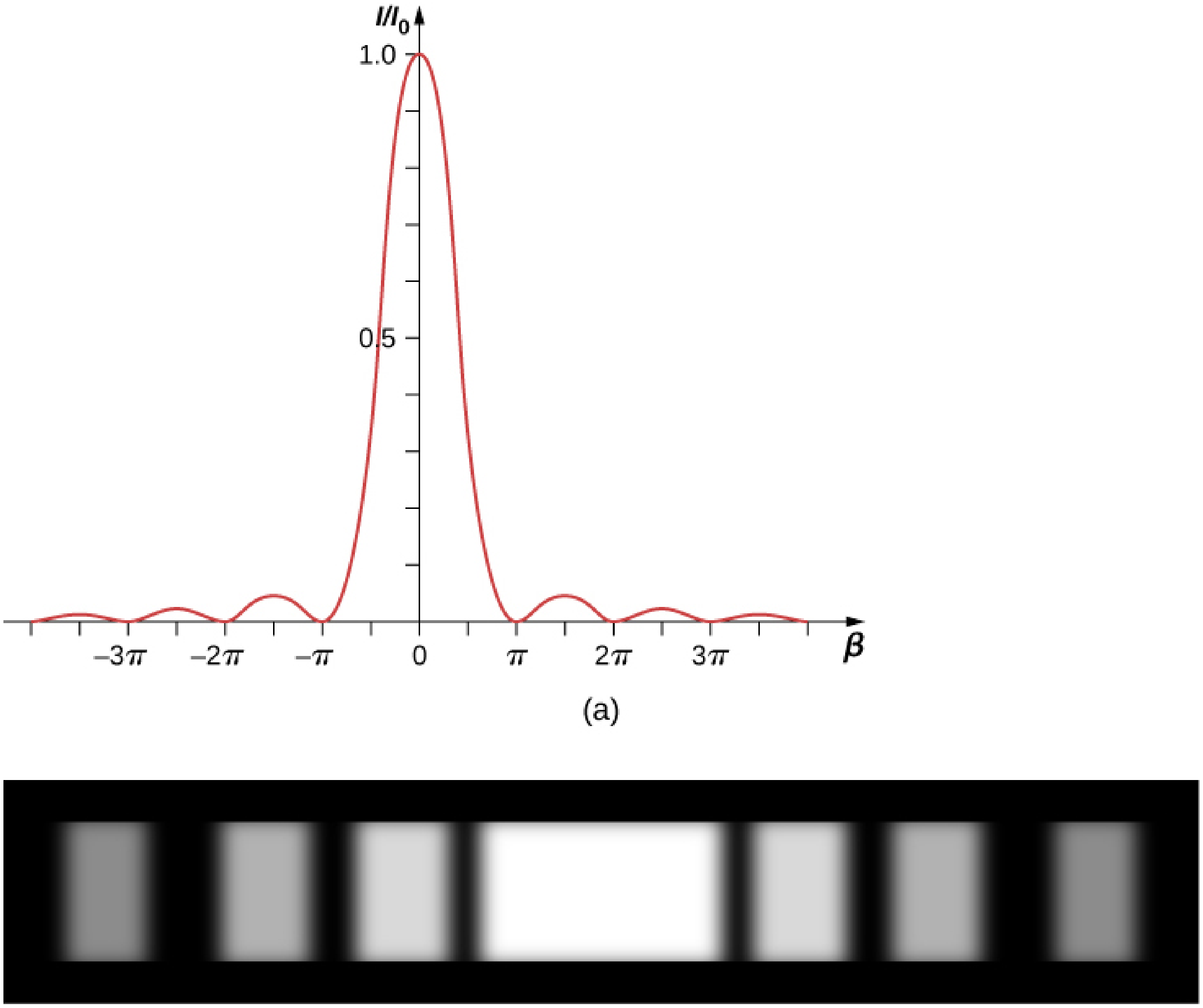
$$\frac{\pi D\sin\theta}{\lambda}=m\pi,$$

or

Equation:

$$D\sin\theta=m\lambda,$$

which is [link].



(a)

(b)

(a) The calculated intensity distribution of a single-slit diffraction pattern. (b) The actual diffraction pattern.

Example:

Intensity in Single-Slit Diffraction

Light of wavelength 550 nm passes through a slit of width $2.00\ \mu$m and produces a diffraction pattern similar to that shown in [link]. (a) Find the locations of the first two minima in terms of the angle from the central maximum and (b) determine the intensity relative to the central maximum at a point halfway between these two minima.

Strategy

The minima are given by [link], $D \sin \theta = m\lambda$. The first two minima are for $m = 1$ and $m = 2$. [link] and [link] can be used to determine the intensity once the angle has been worked out.

Solution

a. Solving [link] for θ gives us $\theta_m = \sin^{-1}(m\lambda/D)$, so that

Equation:

$$\theta_1 = \sin^{-1}\left(\frac{(+1)\left(550\ \times\ 10^{-9}\ \text{m}\right)}{2.00\ \times\ 10^{-6}\ \text{m}}\right) = +16.0^\circ$$

and

Equation:

$$\theta_2 = \sin^{-1}\left(\frac{(+2)\left(550\ \times\ 10^{-9}\ \text{m}\right)}{2.00\ \times\ 10^{-6}\ \text{m}}\right) = +33.4^\circ.$$

b. The halfway point between θ_1 and θ_2 is

Equation:

$$\theta = (\theta_1 + \theta_2)/2 = (16.0^\circ + 33.4^\circ)/2 = 24.7^\circ.$$

[link] gives

Equation:

$$\beta = \frac{\pi D \sin\theta}{\lambda} = \frac{\pi\left(2.00\ \times\ 10^{-6}\,\text{m}\right)\sin\left(24.7^\circ\right)}{\left(550\ \times\ 10^{-9}\,\text{m}\right)} = 1.52\pi\,\text{or}\,4.77\,\text{rad.}$$

From [link], we can calculate

Equation:

$$\frac{I}{I_o} = \left(\frac{\sin\beta}{\beta}\right)^2 = \left(\frac{\sin\ (4.77)}{4.77}\right)^2 = \left(\frac{-0.9985}{4.77}\right)^2 = 0.044.$$

Significance

This position, halfway between two minima, is very close to the location of the maximum, expected near $\beta = 3\pi/2,$ or 1.5π.

Note:

Exercise:

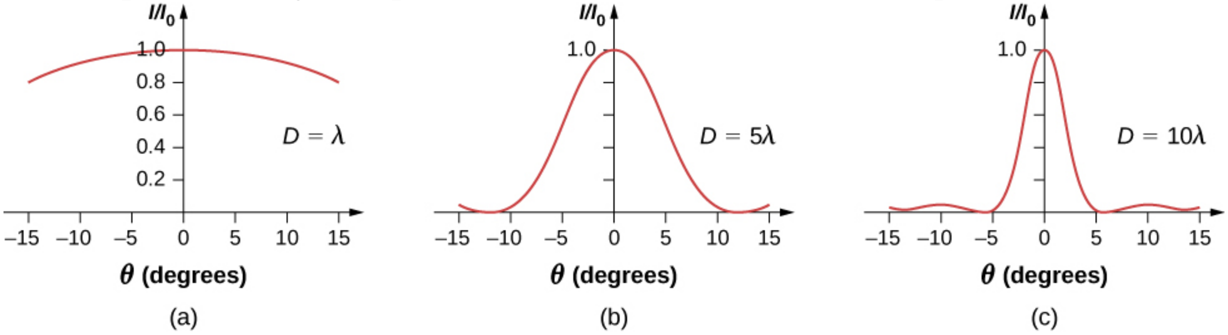
Problem:

Check Your Understanding For the experiment in [link], at what angle from the center is the third maximum and what is its intensity relative to the central maximum?

Solution:

$74.3^\circ, 0.0083I_0$

If the slit width D is varied, the intensity distribution changes, as illustrated in [link]. The central peak is distributed over the region from $\sin\theta = -\lambda/D$ to $\sin\theta = +\lambda/D$. For small θ, this corresponds to an angular width $\Delta\theta \approx 2\lambda/D$. Hence, an increase in the slit width results in a decrease in the **width of the central peak**. For a slit with $D \gg \lambda$, the central peak is very sharp, whereas if $D \approx \lambda$, it becomes quite broad.



Single-slit diffraction patterns for various slit widths. As the slit width D increases from $D = \lambda$ to 5λ and then to 10λ, the width of the central peak decreases as the angles for the first minima decrease as predicted by [link].

Note:
A diffraction experiment in optics can require a lot of preparation but this simulation by Andrew Duffy offers not only a quick set up but also the ability to change the slit width instantly. Run the simulation and select “Single slit.” You can adjust the slit width and see the effect on the diffraction pattern on a screen and as a graph.

Summary

- The intensity pattern for diffraction due to a single slit can be calculated using phasors as
  Equation:

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2,$$

where $\beta = \frac{\phi}{2} = \frac{\pi D \sin\theta}{\lambda}$, D is the slit width, λ is the wavelength, and θ is the angle from the central peak.

Conceptual Questions

Exercise:

Problem:

In [link], the parameter β looks like an angle but is not an angle that you can measure with a protractor in the physical world. Explain what β represents.

Solution:

The parameter $\beta = \phi/2$ is the arc angle shown in the phasor diagram in [link]. The phase difference between the first and last Huygens wavelet across the single slit is 2β and is related to the curvature of the arc that forms the resultant phasor that determines the light intensity.

Problems

Exercise:

Problem:

A single slit of width $3.0\ \mu$m is illuminated by a sodium yellow light of wavelength 589 nm. Find the intensity at a 15° angle to the axis in terms of the intensity of the central maximum.

Exercise:

Problem:

A single slit of width 0.1 mm is illuminated by a mercury light of wavelength 576 nm. Find the intensity at a 10° angle to the axis in terms of the intensity of the central maximum.

Solution:

$I/I_0 = 2.2 \times 10^{-5}$

Exercise:

Problem:

The width of the central peak in a single-slit diffraction pattern is 5.0 mm. The wavelength of the light is 600 nm, and the screen is 2.0 m from the slit. (a) What is the width of the slit? (b) Determine the ratio of the intensity at 4.5 mm from the center of the pattern to the intensity at the center.

Exercise:

Problem:

Consider the single-slit diffraction pattern for $\lambda = 600\,\text{nm}$, $D = 0.025\,\text{mm}$, and $x = 2.0\,\text{m}$. Find the intensity in terms of I_o at $\theta = 0.5^\circ$, 1.0°, 1.5°, 3.0°, and 10.0°.

Solution:

$0.63I_0,\ 0.11I_0,\ 0.0067I_0,\ 0.0062I_0,\ 0.00088I_0$

Glossary

width of the central peak
angle between the minimum for $m = 1$ and $m = -1$

Double-Slit Diffraction

By the end of this section, you will be able to:

- Describe the combined effect of interference and diffraction with two slits, each with finite width
- Determine the relative intensities of interference fringes within a diffraction pattern
- Identify missing orders, if any

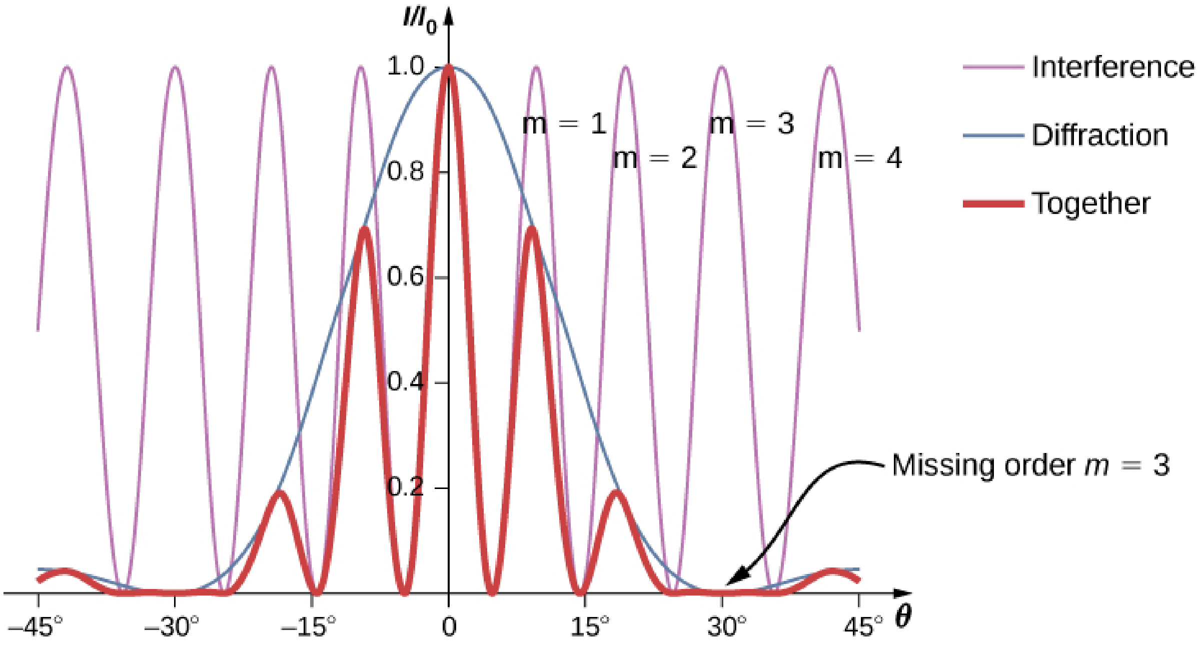
When we studied interference in Young’s double-slit experiment, we ignored the diffraction effect in each slit. We assumed that the slits were so narrow that on the screen you saw only the interference of light from just two point sources. If the slit is smaller than the wavelength, then [link](a) shows that there is just a spreading of light and no peaks or troughs on the screen. Therefore, it was reasonable to leave out the diffraction effect in that chapter. However, if you make the slit wider, [link](b) and (c) show that you cannot ignore diffraction. In this section, we study the complications to the double-slit experiment that arise when you also need to take into account the diffraction effect of each slit.

To calculate the diffraction pattern for two (or any number of) slits, we need to generalize the method we just used for a single slit. That is, across each slit, we place a uniform distribution of point sources that radiate Huygens wavelets, and then we sum the wavelets from all the slits. This gives the intensity at any point on the screen. Although the details of that calculation can be complicated, the final result is quite simple:

Note:
Two-Slit Diffraction Pattern
The diffraction pattern of two slits of width *D* that are separated by a distance *d* is the interference pattern of two point sources separated by *d* multiplied by the diffraction pattern of a slit of width *D*.

In other words, the *locations* of the interference fringes are given by the equation $d\sin\theta = m\lambda$, the same as when we considered the slits to be point sources, but the *intensities* of the fringes are now reduced by diffraction effects, according to [link]. [Note that in the chapter on interference, we wrote $d\sin\theta = m\lambda$ and used the integer *m* to refer to interference fringes. [link] also uses *m*, but this time to refer to diffraction minima. If both equations are used simultaneously, it is good practice to use a different variable (such as *n*) for one of these integers in order to keep them distinct.]

Interference and diffraction effects operate simultaneously and generally produce minima at different angles. This gives rise to a complicated pattern on the screen, in which some of the maxima of interference from the two slits are missing if the maximum of the interference is in the same direction as the minimum of the diffraction. We refer to such a missing peak as a **missing order**. One example of a diffraction pattern on the screen is shown in [link]. The solid line with multiple peaks of various heights is the intensity observed on the screen. It is a product of the interference pattern of waves from separate slits and the diffraction of waves from within one slit.



Diffraction from a double slit. The purple line with peaks of the same height are from the interference of the waves from two slits; the blue line with one big hump in the middle is the diffraction of waves from within one slit; and the thick red line is the product of the two, which is the pattern observed on the screen. The plot shows the expected result for a slit width $D = 2\lambda$ and slit separation $d = 6\lambda$. The maximum of $m = \pm 3$ order for the interference is missing because the minimum of the diffraction occurs in the same direction.

Example:

Intensity of the Fringes

[link] shows that the intensity of the fringe for $m = 3$ is zero, but what about the other fringes? Calculate the intensity for the fringe at $m = 1$ relative to I_0, the intensity of the central peak.

Strategy

Determine the angle for the double-slit interference fringe, using the equation from Interference, then determine the relative intensity in that direction due to diffraction by using [link].

Solution

From the chapter on interference, we know that the bright interference fringes occur at $d \sin \theta = m\lambda$, or

Equation:

$$\sin\theta = \frac{m\lambda}{d}.$$

From [link],

Equation:

$$I = I_0\left(\frac{\sin\beta}{\beta}\right)^2, \text{ where } \beta = \frac{\phi}{2} = \frac{\pi D\sin\theta}{\lambda}.$$

Substituting from above,

Equation:

$$\beta = \frac{\pi D\sin\theta}{\lambda} = \frac{\pi D}{\lambda}\cdot\frac{m\lambda}{d} = \frac{m\pi D}{d}.$$

For $D = 2\lambda$, $d = 6\lambda$, and $m = 1$,

Equation:

$$\beta = \frac{(1)\pi(2\lambda)}{(6\lambda)} = \frac{\pi}{3}.$$

Then, the intensity is

Equation:

$$I = I_0\left(\frac{\sin\beta}{\beta}\right)^2 = I_0\left(\frac{\sin\ (\pi/3)}{\pi/3}\right)^2 = 0.684I_0.$$

Significance

Note that this approach is relatively straightforward and gives a result that is almost exactly the same as the more complicated analysis using phasors to work out the intensity values of the double-slit interference (thin line in [link]). The phasor approach accounts for the downward slope in the diffraction intensity (blue line) so that the peak *near* $m = 1$ occurs at a value of θ ever so slightly smaller than we have shown here.

Example:

Two-Slit Diffraction

Suppose that in Young’s experiment, slits of width 0.020 mm are separated by 0.20 mm. If the slits are illuminated by monochromatic light of wavelength 500 nm, how many bright fringes are observed in the central peak of the diffraction pattern?

Solution

From [link], the angular position of the first diffraction minimum is $\theta \approx \sin\theta = \frac{\lambda}{D} = \frac{5.0\times10^{-7}\,\text{m}}{2.0\times10^{-5}\,\text{m}} = 2.5\ \times\ 10^{-2}\,\text{rad}$.
Using $d\sin\theta = m\lambda$ for $\theta = 2.5\ \times\ 10^{-2}\,\text{rad}$, we find

Equation:

$$m = \frac{d\sin\theta}{\lambda} = \frac{(0.20\,\text{mm})\left(2.5\ \times\ 10^{-2}\,\text{rad}\right)}{\left(5.0\ \times\ 10^{-7}\,\text{m}\right)} = 10,$$

which is the maximum interference order that fits inside the central peak. We note that $m = \pm10$ are missing orders as θ matches exactly. Accordingly, we observe bright fringes for

Equation:

$$m = -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, +1, +2, +3, +4, +5, +6, +7, +8,\ \text{and}\ +9$$

for a total of 19 bright fringes.

Note:

Exercise:

Problem:

Check Your Understanding For the experiment in [link], show that $m = 20$ is also a missing order.

Solution:

From $d\sin\theta = m\lambda$, the interference maximum occurs at 2.87° for $m = 20$. From [link], this is also the angle for the second diffraction minimum. (*Note:* Both equations use the index *m* but they refer to separate phenomena.)

Note:

Explore the effects of double-slit diffraction. In this simulation written by Fu-Kwun Hwang, select $N = 2$ using the slider and see what happens when you control the slit width, slit separation and the wavelength. Can you make an order go “missing?”

Summary

- With real slits with finite widths, the effects of interference and diffraction operate simultaneously to form a complicated intensity pattern.

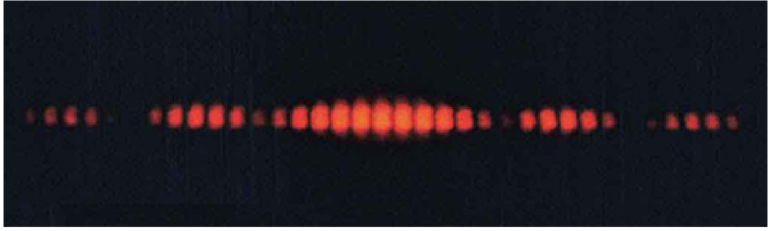
- Relative intensities of interference fringes within a diffraction pattern can be determined.
- Missing orders occur when an interference maximum and a diffraction minimum are located together.

Conceptual Questions

Exercise:

Problem:

Shown below is the central part of the interference pattern for a pure wavelength of red light projected onto a double slit. The pattern is actually a combination of single- and double-slit interference. Note that the bright spots are evenly spaced. Is this a double- or single-slit characteristic? Note that some of the bright spots are dim on either side of the center. Is this a single- or double-slit characteristic? Which is smaller, the slit width or the separation between slits? Explain your responses.



(credit: PASCO)

Problems

Exercise:

Problem:

Two slits of width $2\ \mu\text{m}$, each in an opaque material, are separated by a center-to-center distance of $6\ \mu\text{m}$. A monochromatic light of wavelength 450 nm is incident on the double-slit. One finds a combined interference and diffraction pattern on the screen.

(a) How many peaks of the interference will be observed in the central maximum of the diffraction pattern?

(b) How many peaks of the interference will be observed if the slit width is doubled while keeping the distance between the slits same?

(c) How many peaks of interference will be observed if the slits are separated by twice the distance, that is, $12\ \mu\text{m}$, while keeping the widths of the slits same?

(d) What will happen in (a) if instead of 450-nm light another light of wavelength 680 nm is used?

(e) What is the value of the ratio of the intensity of the central peak to the intensity of the next bright peak in (a)?

(f) Does this ratio depend on the wavelength of the light?

(g) Does this ratio depend on the width or separation of the slits?

Exercise:

Problem:

A double slit produces a diffraction pattern that is a combination of single- and double-slit interference. Find the ratio of the width of the slits to the separation between them, if the first minimum of the single-slit pattern falls on the fifth maximum of the double-slit pattern. (This will greatly reduce the intensity of the fifth maximum.)

Solution:

0.200

Exercise:

Problem:

For a double-slit configuration where the slit separation is four times the slit width, how many interference fringes lie in the central peak of the diffraction pattern?

Exercise:

Problem:

Light of wavelength 500 nm falls normally on 50 slits that are $2.5 \times 10^{-3}\,\text{mm}$ wide and spaced $5.0 \times 10^{-3}\,\text{mm}$ apart. How many interference fringes lie in the central peak of the diffraction pattern?

Solution:

3

Exercise:

Problem:

A monochromatic light of wavelength 589 nm incident on a double slit with slit width $2.5\,\mu\text{m}$ and unknown separation results in a diffraction pattern containing nine interference peaks inside the central maximum. Find the separation of the slits.

Exercise:

Problem:

When a monochromatic light of wavelength 430 nm incident on a double slit of slit separation $5\ \mu\mathrm{m}$, there are 11 interference fringes in its central maximum. How many interference fringes will be in the central maximum of a light of the same wavelength and slit widths, but a new slit separation of $4\ \mu\mathrm{m}$?

Solution:

9

Exercise:

Problem:

Determine the intensities of two interference peaks other than the central peak in the central maximum of the diffraction, if possible, when a light of wavelength 628 nm is incident on a double slit of width 500 nm and separation 1500 nm. Use the intensity of the central spot to be $1\ \mathrm{mW/cm^2}$.

Glossary

missing order	interference maximum that is not seen because it coincides with a diffraction minimum
two-slit diffraction pattern	diffraction pattern of two slits of width *D* that are separated by a distance *d* is the interference pattern of two point sources separated by *d* multiplied by the diffraction pattern of a slit of width *D*

Diffraction Gratings

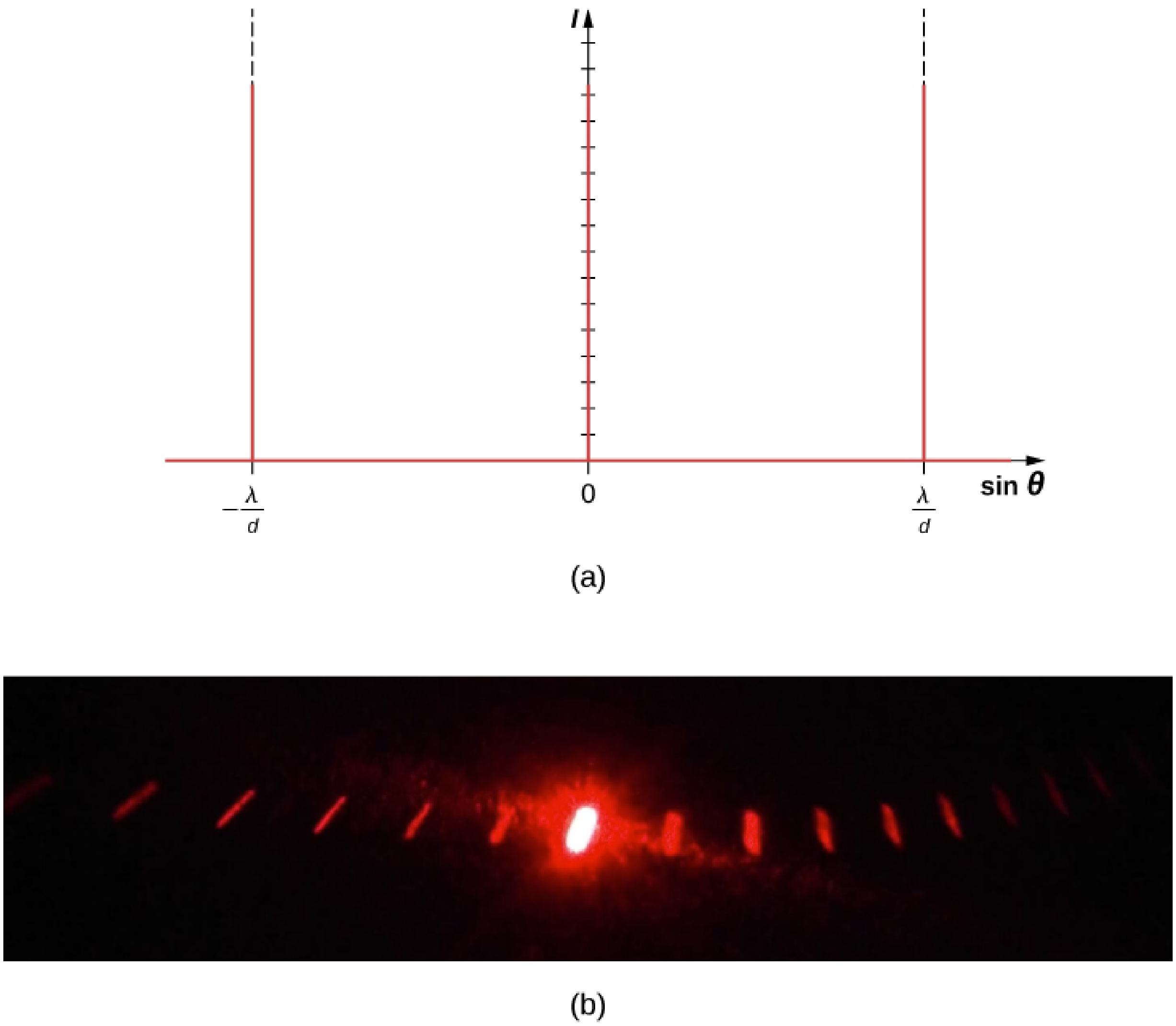
By the end of this section, you will be able to:

- Discuss the pattern obtained from diffraction gratings
- Explain diffraction grating effects

Analyzing the interference of light passing through two slits lays out the theoretical framework of interference and gives us a historical insight into Thomas Young’s experiments. However, most modern-day applications of slit interference use not just two slits but many, approaching infinity for practical purposes. The key optical element is called a diffraction grating, an important tool in optical analysis.

Diffraction Gratings: An Infinite Number of Slits

The analysis of multi-slit interference in Interference allows us to consider what happens when the number of slits N approaches infinity. Recall that $N-2$ secondary maxima appear between the principal maxima. We can see there will be an infinite number of secondary maxima that appear, and an infinite number of dark fringes between them. This makes the spacing between the fringes, and therefore the width of the maxima, infinitesimally small. Furthermore, because the intensity of the secondary maxima is proportional to $1/N^2$, it approaches zero so that the secondary maxima are no longer seen. What remains are only the principal maxima, now very bright and very narrow ([link]).

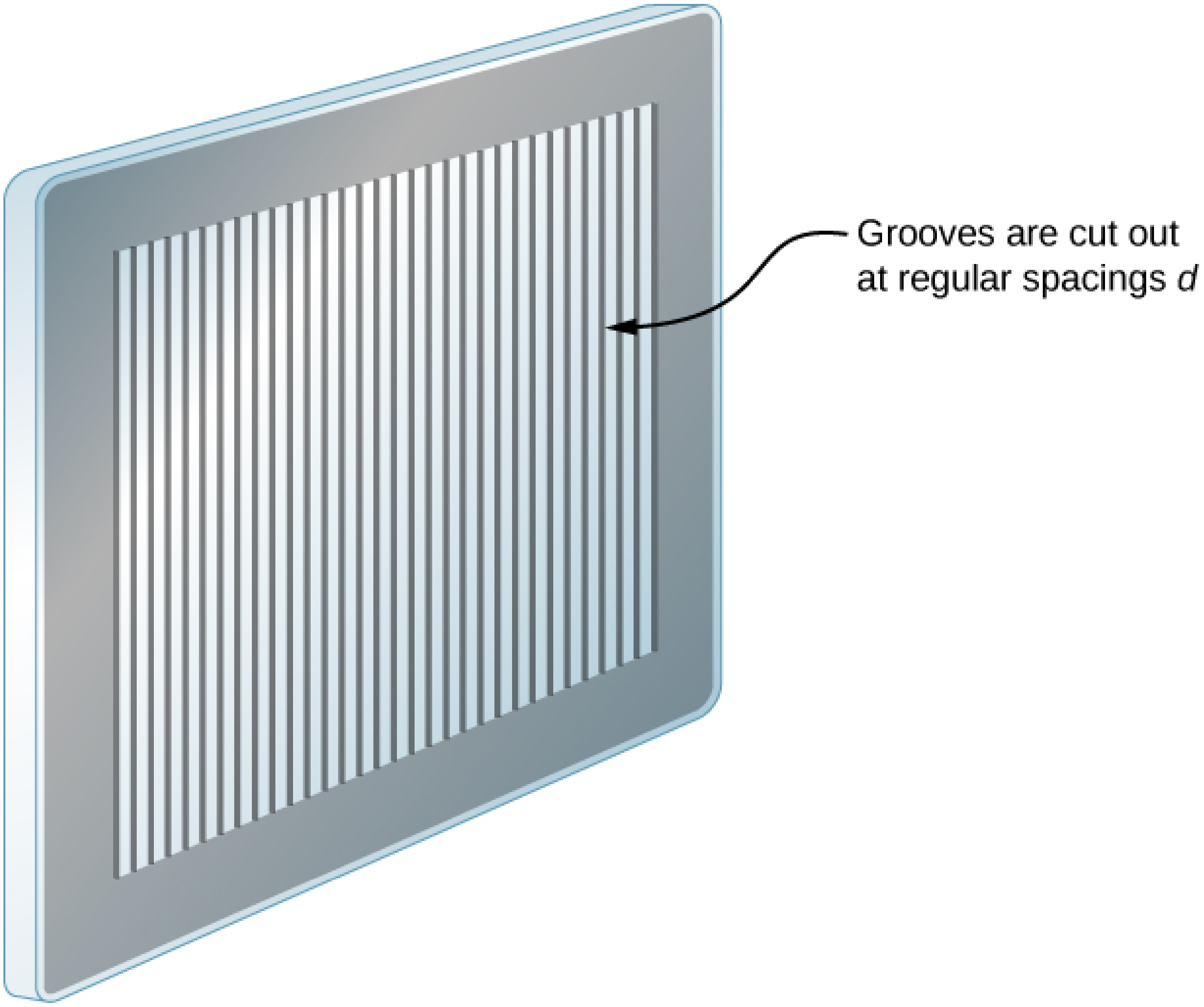


(a)

(b)

(a) Intensity of light transmitted through a large number of slits. When *N* approaches infinity, only the principal maxima remain as very bright and very narrow lines. (b) A laser beam passed through a diffraction grating. (credit b: modification of work by Sebastian Stapelberg)

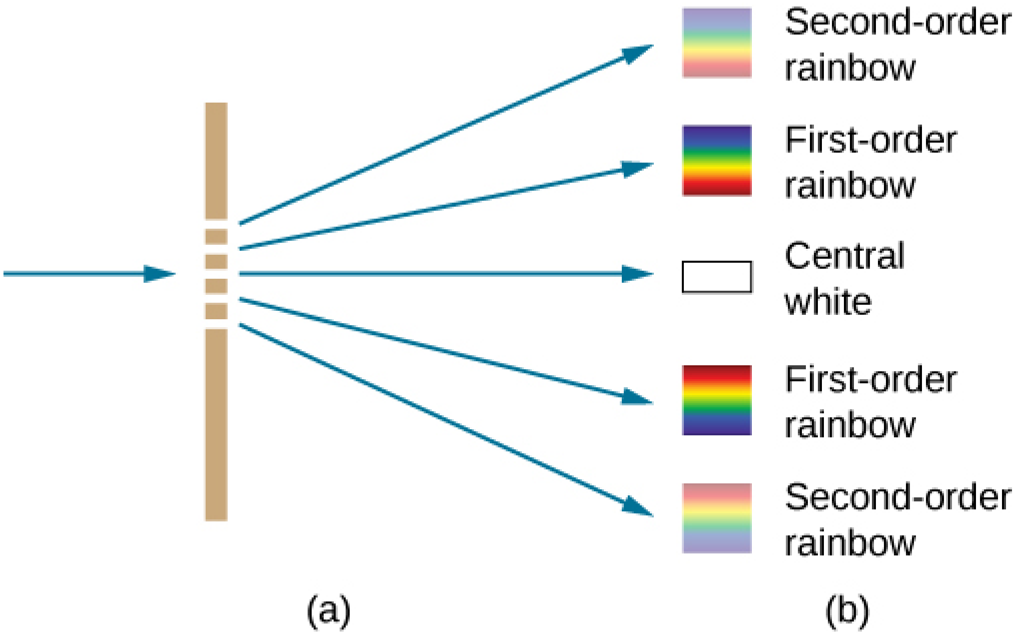
In reality, the number of slits is not infinite, but it can be very large—large enough to produce the equivalent effect. A prime example is an optical element called a **diffraction grating**. A diffraction grating can be manufactured by carving glass with a sharp tool in a large number of precisely positioned parallel lines, with untouched regions acting like slits ([link]). This type of grating can be photographically mass produced rather cheaply. Because there can be over 1000 lines per millimeter across the grating, when a section as small as a few millimeters is illuminated by an incoming ray, the number of illuminated slits is effectively infinite, providing for very sharp principal maxima.



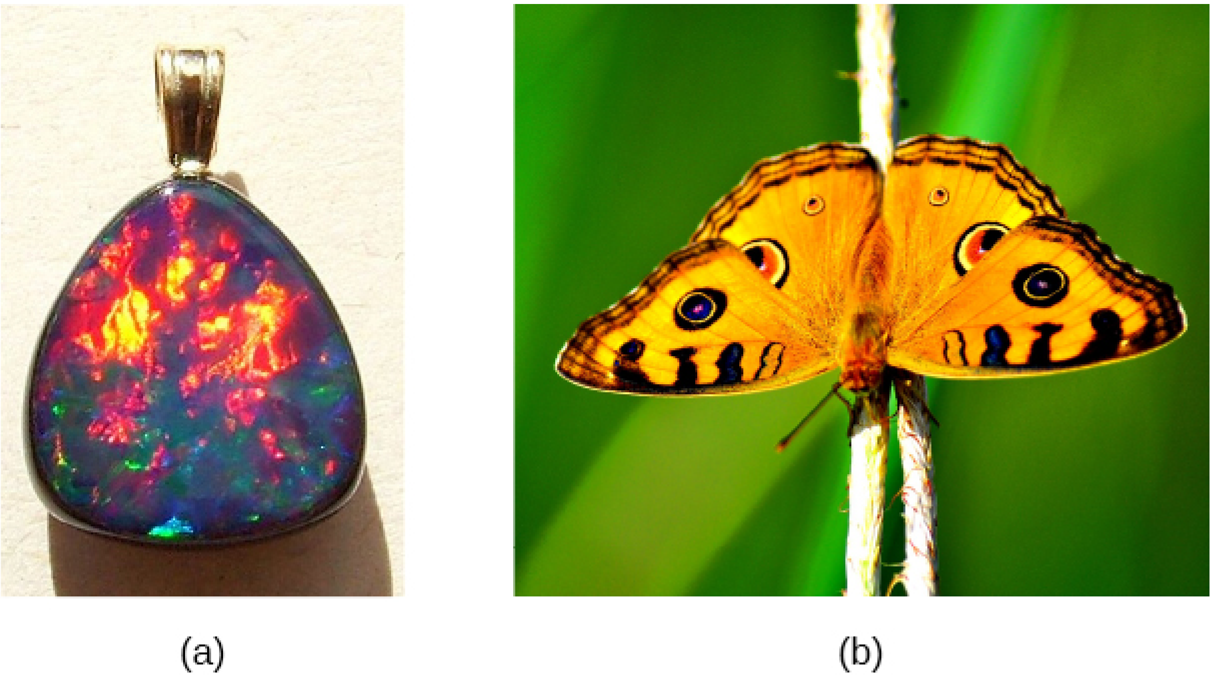
A diffraction grating can be manufactured by carving glass with a sharp tool in a large number of precisely positioned parallel lines.

Diffraction gratings work both for transmission of light, as in [link], and for reflection of light, as on butterfly wings and the Australian opal in [link].

Natural diffraction gratings also occur in the feathers of certain birds such as the hummingbird. Tiny, finger-like structures in regular patterns act as reflection gratings, producing constructive interference that gives the feathers colors not solely due to their pigmentation. This is called iridescence.



(a) Light passing through a diffraction grating is diffracted in a pattern similar to a double slit, with bright regions at various angles. (b) The pattern obtained for white light incident on a grating. The central maximum is white, and the higher-order maxima disperse white light into a rainbow of colors.



(a) This Australian opal and (b) butterfly wings have rows of reflectors that act like reflection gratings, reflecting different colors at different angles. (credit a: modification of work by "Opals-On-Black"/Flickr; credit b: modification of work by “whologwhy”/Flickr)

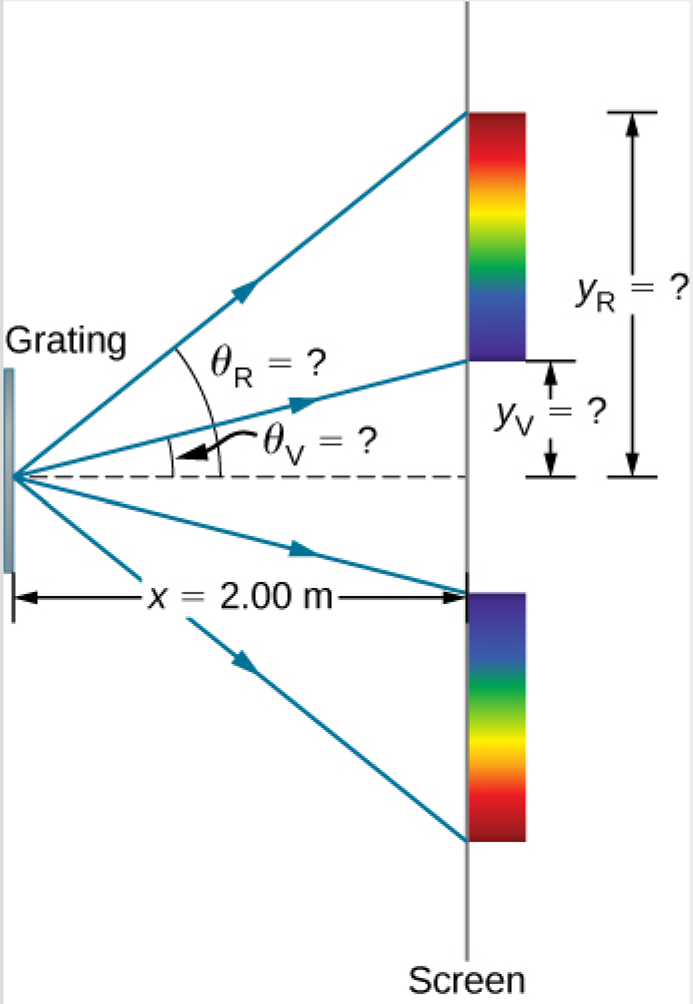
Applications of Diffraction Gratings

Where are diffraction gratings used in applications? Diffraction gratings are commonly used for spectroscopic dispersion and analysis of light. What makes them particularly useful is the fact that they form a sharper pattern than double slits do. That is, their bright fringes are narrower and brighter while their dark regions are darker. Diffraction gratings are key components of monochromators used, for example, in optical imaging of particular wavelengths from biological or medical samples. A diffraction grating can be chosen to specifically analyze a wavelength emitted by molecules in diseased cells in a biopsy sample or to help excite strategic molecules in the sample with a selected wavelength of light. Another vital use is in optical fiber technologies where fibers are designed to provide optimum performance at specific wavelengths. A range of diffraction gratings are available for selecting wavelengths for such use.

Example:

Calculating Typical Diffraction Grating Effects

Diffraction gratings with 10,000 lines per centimeter are readily available. Suppose you have one, and you send a beam of white light through it to a screen 2.00 m away. (a) Find the angles for the first-order diffraction of the shortest and longest wavelengths of visible light (380 and 760 nm, respectively). (b) What is the distance between the ends of the rainbow of visible light produced on the screen for first-order interference? (See [link].)



(a) The diffraction grating considered in this example produces a rainbow of colors on a screen a distance $x = 2.00\,\text{m}$ from the grating. The distances along the screen are measured perpendicular to the *x*-direction. In other words, the rainbow pattern extends out of the page.
(b) In a bird's-eye view, the rainbow pattern can be seen on a table where the equipment is placed.

Strategy

Once a value for the diffraction grating's slit spacing *d* has been determined, the angles for the sharp lines can be found using the equation

Equation:

$$d\sin\theta = m\lambda\,\text{for}\,m = 0,\ \pm1,\ \pm2,\ \ldots.$$

Since there are 10,000 lines per centimeter, each line is separated by 1/10,000 of a centimeter. Once we know the angles, we an find the distances along the screen by using simple trigonometry.

Solution

a. The distance between slits is $d = (1\,\text{cm})/10,000 = 1.00\,\times\,10^{-4}\,\text{cm}$ or $1.00\,\times\,10^{-6}\,\text{m}$. Let us call the two angles θ_V for violet (380 nm) and θ_R for red (760 nm). Solving the equation $d\sin\theta_\text{V} = m\lambda$ for $\sin\theta_\text{V}$,

Equation:

$$\sin\theta_\text{V} = \frac{m\lambda_\text{V}}{d},$$

where $m = 1$ for the first-order and $\lambda_\text{V} = 380\,\text{nm} = 3.80\,\times\,10^{-7}\,\text{m}$. Substituting these values gives

Equation:

$$\sin\theta_{\text{V}} = \frac{3.80\ \times\ 10^{-7}\,\text{m}}{1.00\ \times\ 10^{-6}\,\text{m}} = 0.380.$$

Thus the angle θ_{V} is
Equation:

$$\theta_{\text{V}} = \sin^{-1} 0.380 = 22.33^{\circ}.$$

Similarly,
Equation:

$$\sin\theta_{\text{R}} = \frac{7.60\ \times\ 10^{-7}\,\text{m}}{1.00\ \times\ 10^{-6}\,\text{m}} = 0.760.$$

Thus the angle θ_{R} is
Equation:

$$\theta_{\text{R}} = \sin^{-1} 0.760 = 49.46^{\circ}.$$

Notice that in both equations, we reported the results of these intermediate calculations to four significant figures to use with the calculation in part (b).

b. The distances on the secreen are labeled y_{V} and y_{R} in [link]. Notice that $\tan\theta = y/x$. We can solve for y_{V} and y_{R}. That is,
Equation:

$$y_{\text{V}} = x\tan\theta_{V} = (2.00\,\text{m})(\tan 22.33^{\circ}) = 0.815\,\text{m}$$

and
Equation:

$$y_{\text{R}} = x\tan\theta_{R} = (2.00\,\text{m})(\tan 49.46^{\circ}) = 2.338\,\text{m}.$$

The distance between them is therefore

Equation:

$$y_{\mathrm{R}} - y_{\mathrm{V}} = 1.523\,\mathrm{m}.$$

Significance

The large distance between the red and violet ends of the rainbow produced from the white light indicates the potential this diffraction grating has as a spectroscopic tool. The more it can spread out the wavelengths (greater dispersion), the more detail can be seen in a spectrum. This depends on the quality of the diffraction grating—it must be very precisely made in addition to having closely spaced lines.

Note:

Exercise:

Problem:

Check Your Understanding If the line spacing of a diffraction grating *d* is not precisely known, we can use a light source with a well-determined wavelength to measure it. Suppose the first-order constructive fringe of the H_{β} emission line of hydrogen $(\lambda = 656.3\,\mathrm{nm})$ is measured at 11.36° using a spectrometer with a diffraction grating. What is the line spacing of this grating?

Solution:

3.332×10^{-6} m or 300 lines per millimeter

Note:

Take the same simulation we used for double-slit diffraction and try increasing the number of slits from $N = 2$ to $N = 3, 4, 5\text{…}$. The primary peaks become sharper, and the secondary peaks become less and less pronounced. By the time you reach the maximum number of $N = 20$, the system is behaving much like a diffraction grating.

Summary

- A diffraction grating consists of a large number of evenly spaced parallel slits that produce an interference pattern similar to but sharper than that of a double slit.
- Constructive interference occurs when $d \sin \theta = m\lambda$ for $m = 0,\ \pm1,\ \pm2,\ \ldots$, where d is the distance between the slits, θ is the angle relative to the incident direction, and m is the order of the interference.

Problems

Exercise:

Problem:

A diffraction grating has 2000 lines per centimeter. At what angle will the first-order maximum be for 520-nm-wavelength green light?

Solution:

5.97°

Exercise:

Problem:

Find the angle for the third-order maximum for 580-nm-wavelength yellow light falling on a difraction grating having 1500 lines per centimeter.

Exercise:

Problem:

How many lines per centimeter are there on a diffraction grating that gives a first-order maximum for 470-nm blue light at an angle of 25.0° ?

Solution:

8.99×10^3

Exercise:

Problem:

What is the distance between lines on a diffraction grating that produces a second-order maximum for 760-nm red light at an angle of 60.0° ?

Exercise:

Problem:

Calculate the wavelength of light that has its second-order maximum at 45.0° when falling on a diffraction grating that has 5000 lines per centimeter.

Solution:

707 nm

Exercise:

Problem:

An electric current through hydrogen gas produces several distinct wavelengths of visible light. What are the wavelengths of the hydrogen spectrum, if they form first-order maxima at angles $24.2^\circ,\ 25.7^\circ,\ 29.1^\circ$, and 41.0° when projected on a diffraction grating having 10,000 lines per centimeter?

Exercise:

Problem:

(a) What do the four angles in the preceding problem become if a 5000-line per centimeter diffraction grating is used? (b) Using this grating, what would the angles be for the second-order maxima? (c) Discuss the relationship between integral reductions in lines per centimeter and the new angles of various order maxima.

Solution:

a. 11.8°, 12.5°, 14.1°, 19.2°; b. 24.2°, 25.7°, 29.1°, 41.0°; c. Decreasing the number of lines per centimeter by a factor of *x* means that the angle for the *x*-order maximum is the same as the original angle for the first-order maximum.

Exercise:

Problem:

What is the spacing between structures in a feather that acts as a reflection grating, giving that they produce a first-order maximum for 525-nm light at a 30.0° angle?

Exercise:

Problem:

An opal such as that shown in [link] acts like a reflection grating with rows separated by about $8\ \mu\text{m}$. If the opal is illuminated normally, (a) at what angle will red light be seen and (b) at what angle will blue light be seen?

Solution:

a. using $\lambda = 700\ \text{nm}$, $\theta = 5.0^\circ$; b. using $\lambda = 460\ \text{nm}$, $\theta = 3.3^\circ$

Exercise:

Problem:

At what angle does a diffraction grating produce a second-order maximum for light having a first-order maximum at 20.0° ?

Exercise:

Problem:

(a) Find the maximum number of lines per centimeter a diffraction grating can have and produce a maximum for the smallest wavelength of visible light. (b) Would such a grating be useful for ultraviolet spectra? (c) For infrared spectra?

Solution:

a. 26,300 lines/cm; b. yes; c. no

Exercise:

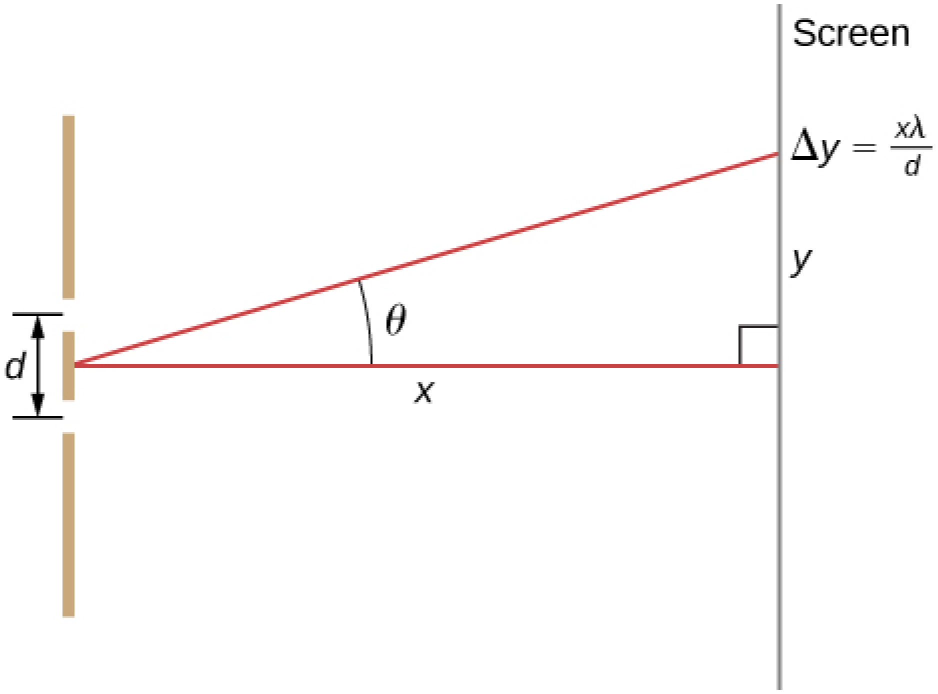
Problem:

(a) Show that a 30,000 line per centimeter grating will not produce a maximum for visible light. (b) What is the longest wavelength for which it does produce a first-order maximum? (c) What is the greatest number of line per centimeter a diffraction grating can have and produce a complete second-order spectrum for visible light?

Exercise:

Problem:

The analysis shown below also applies to diffraction gratings with lines separated by a distance *d*. What is the distance between fringes produced by a diffraction grating having 125 lines per centimeter for 600-nm light, if the screen is 1.50 m away? (*Hin*t*:* The distance between adjacent fringes is $\Delta y = x\lambda/d$, assuming the slit separation *d* is comparable to λ.)



Solution:

$1.13 \times 10^{-2}\,\text{m}$

Glossary

diffraction grating
: large number of evenly spaced parallel slits

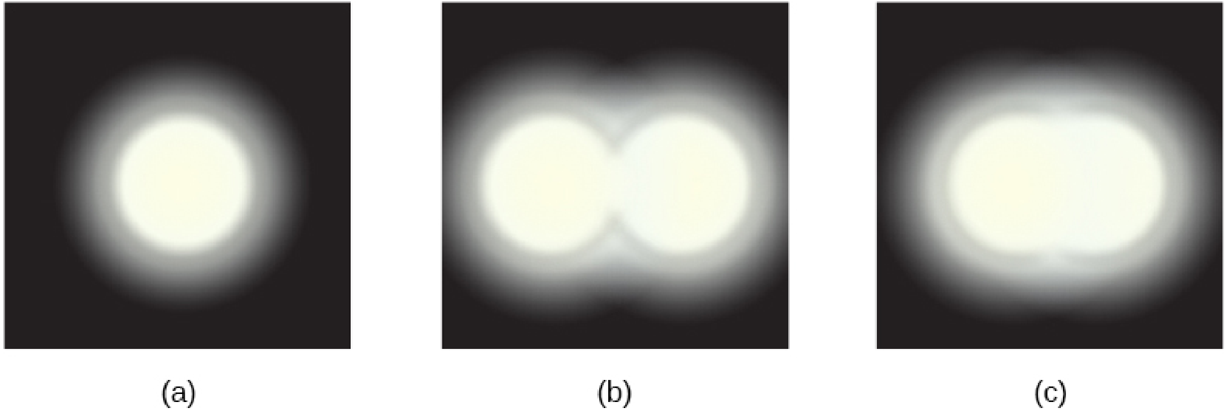
Circular Apertures and Resolution

By the end of this section, you will be able to:

- Describe the diffraction limit on resolution
- Describe the diffraction limit on beam propagation

Light diffracts as it moves through space, bending around obstacles, interfering constructively and destructively. This can be used as a spectroscopic tool—a diffraction grating disperses light according to wavelength, for example, and is used to produce spectra—but diffraction also limits the detail we can obtain in images.

[link](a) shows the effect of passing light through a small circular aperture. Instead of a bright spot with sharp edges, we obtain a spot with a fuzzy edge surrounded by circles of light. This pattern is caused by diffraction, similar to that produced by a single slit. Light from different parts of the circular aperture interferes constructively and destructively. The effect is most noticeable when the aperture is small, but the effect is there for large apertures as well.



(a) Monochromatic light passed through a small circular aperture produces this diffraction pattern. (b) Two point-light sources that are close to one another produce overlapping images because of diffraction. (c) If the sources are closer together, they cannot be distinguished or resolved.

How does diffraction affect the detail that can be observed when light passes through an aperture? [link](b) shows the diffraction pattern produced by two point-light sources that are close to one another. The pattern is similar to that for a single point source, and it is still possible to tell that there are two light sources rather than one. If they are closer together, as in [link](c), we cannot distinguish them, thus limiting the detail or **resolution** we can obtain. This limit is an inescapable consequence of the wave nature of light.

Diffraction limits the resolution in many situations. The acuity of our vision is limited because light passes through the pupil, which is the circular aperture of the eye. Be aware that the diffraction-like spreading of light is due to the limited diameter of a light beam, not the interaction with an aperture. Thus, light passing through a lens with a diameter *D* shows this effect and spreads, blurring the image, just as light passing through an aperture of diameter *D* does. Thus, diffraction limits the resolution of any system having a lens or mirror. Telescopes are also limited by diffraction, because of the finite diameter *D* of the primary mirror.

Just what is the limit? To answer that question, consider the diffraction pattern for a circular aperture, which has a central maximum that is wider and brighter than the maxima surrounding it (similar to a slit) ([link](a)). It can be shown that, for a circular aperture of diameter *D*, the first minimum in the diffraction pattern occurs at $\theta = 1.22\lambda/D$ (providing the aperture is large compared with the wavelength of light, which is the case for most optical instruments). The accepted criterion for determining the **diffraction limit** to resolution based on this angle is known as the **Rayleigh criterion**, which was developed by Lord Rayleigh in the nineteenth century.

Note:

Rayleigh Criterion

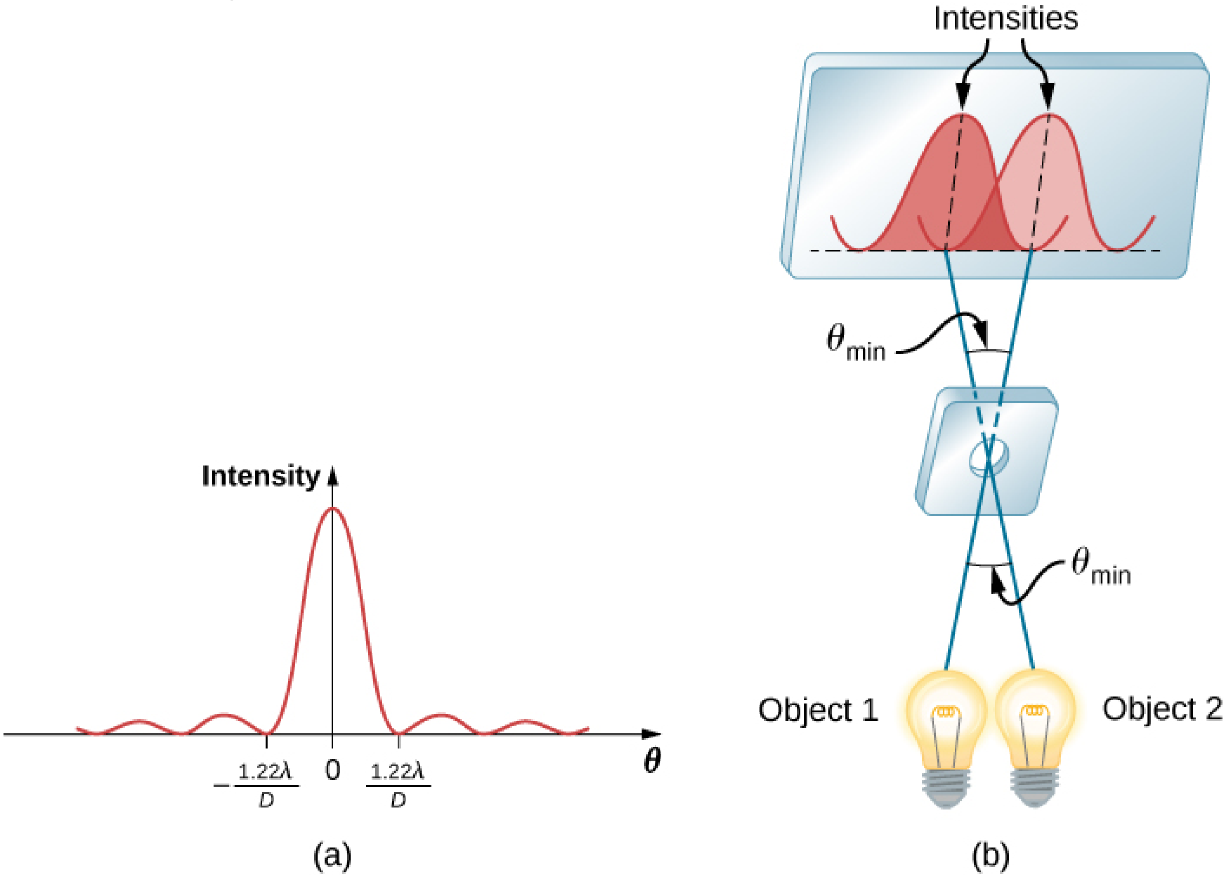
The diffraction limit to resolution states that two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other ([link](b)).

The first minimum is at an angle of $\theta = 1.22\lambda/D$, so that two point objects are just resolvable if they are separated by the angle

Note:
Equation:

$$\theta = 1.22\frac{\lambda}{D}$$

where λ is the wavelength of light (or other electromagnetic radiation) and D is the diameter of the aperture, lens, mirror, etc., with which the two objects are observed. In this expression, θ has units of radians. This angle is also commonly known as the diffraction limit.



(a) Graph of intensity of the diffraction pattern for a circular aperture. Note that, similar to a single slit, the central maximum is wider and brighter than those to the sides. (b) Two point objects produce overlapping diffraction patterns. Shown here is the Rayleigh criterion for being just resolvable. The central maximum of one pattern lies on the first minimum of the other.

All attempts to observe the size and shape of objects are limited by the wavelength of the probe. Even the small wavelength of light prohibits exact precision. When extremely small wavelength probes are used, as with an electron microscope, the system is disturbed, still limiting our knowledge. Heisenberg’s uncertainty principle asserts that this limit is fundamental and inescapable, as we shall see in the chapter on quantum mechanics.

Example:

Calculating Diffraction Limits of the Hubble Space Telescope

The primary mirror of the orbiting Hubble Space Telescope has a diameter of 2.40 m. Being in orbit, this telescope avoids the degrading effects of atmospheric distortion on its resolution. (a) What is the angle between two just-resolvable point light sources (perhaps two stars)? Assume an average light wavelength of 550 nm. (b) If these two stars are at a distance of 2 million light-years, which is the distance of the Andromeda Galaxy, how close together can they be and still be resolved? (A light-year, or ly, is the distance light travels in 1 year.)

Strategy

The Rayleigh criterion stated in [link], $\theta = 1.22\lambda/D$, gives the smallest possible angle θ between point sources, or the best obtainable resolution. Once this angle is known, we can calculate the distance between the stars, since we are given how far away they are.

Solution

a. The Rayleigh criterion for the minimum resolvable angle is
   Equation:

$$\theta = 1.22\frac{\lambda}{D}.$$

Entering known values gives

Equation:

$$\theta = 1.22\frac{550 \times 10^{-9}\,\text{m}}{2.40\,\text{m}} = 2.80 \times 10^{-7}\,\text{rad}.$$

b. The distance *s* between two objects a distance *r* away and separated by an angle θ is $s = r\theta$.

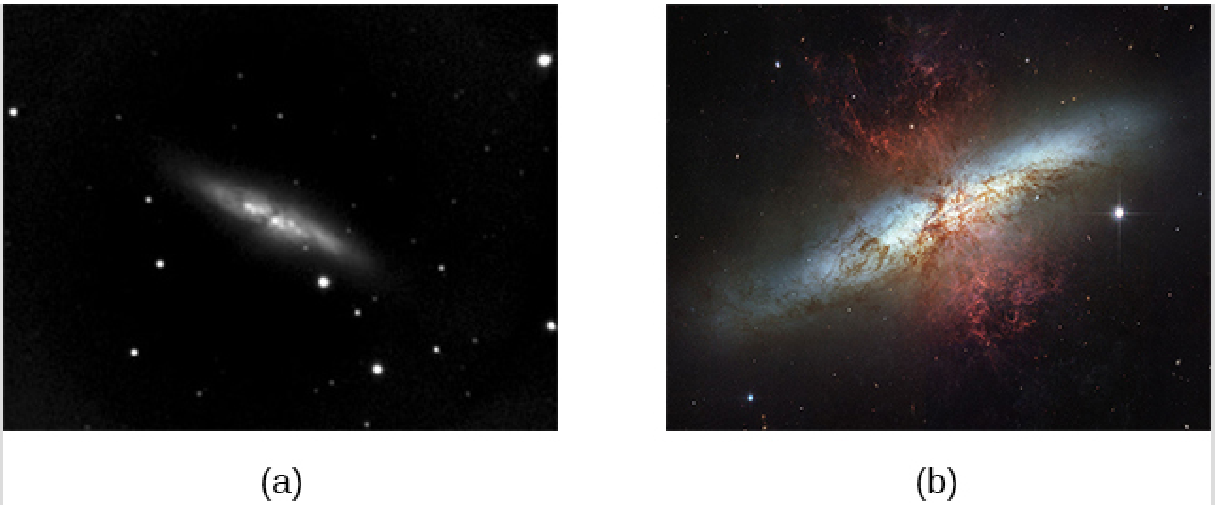
Substituting known values gives

Equation:

$$s = \left(2.0 \times 10^{6}\,\text{ly}\right)\left(2.80 \times 10^{-7}\,\text{rad}\right) = 0.56\,\text{ly}.$$

Significance

The angle found in part (a) is extraordinarily small (less than 1/50,000 of a degree), because the primary mirror is so large compared with the wavelength of light. As noticed, diffraction effects are most noticeable when light interacts with objects having sizes on the order of the wavelength of light. However, the effect is still there, and there is a diffraction limit to what is observable. The actual resolution of the Hubble Telescope is not quite as good as that found here. As with all instruments, there are other effects, such as nonuniformities in mirrors or aberrations in lenses that further limit resolution. However, [link] gives an indication of the extent of the detail observable with the Hubble because of its size and quality, and especially because it is above Earth’s atmosphere.


(a) (b)

These two photographs of the M82 Galaxy give an idea of the observable detail using (a) a ground-based telescope and (b) the Hubble Space Telescope. (credit a: modification of work by “Ricnun”/Wikimedia Commons; credit b: modification of work by NASA, ESA, and The Hubble Heritage Team (STScI/AURA))

The answer in part (b) indicates that two stars separated by about half a light-year can be resolved. The average distance between stars in a galaxy is on the order of five light-years in the outer parts and about one light-year near the galactic center. Therefore, the Hubble can resolve most of the individual stars in Andromeda Galaxy, even though it lies at such a huge distance that its light takes 2 million years to reach us. [link] shows another mirror used to observe radio waves from outer space.

A 305-m-diameter paraboloid at Arecibo in Puerto Rico is lined with reflective material, making it into a radio telescope. It is the largest curved focusing dish in the world. Although *D* for Arecibo is much larger than for the Hubble Telescope, it detects radiation of a much longer wavelength and its diffraction limit is significantly poorer than Hubble’s. The Arecibo telescope is still very useful, because important information is carried by radio waves that is not carried by visible light. (credit: Jeff Hitchcock)

Note:

Exercise:

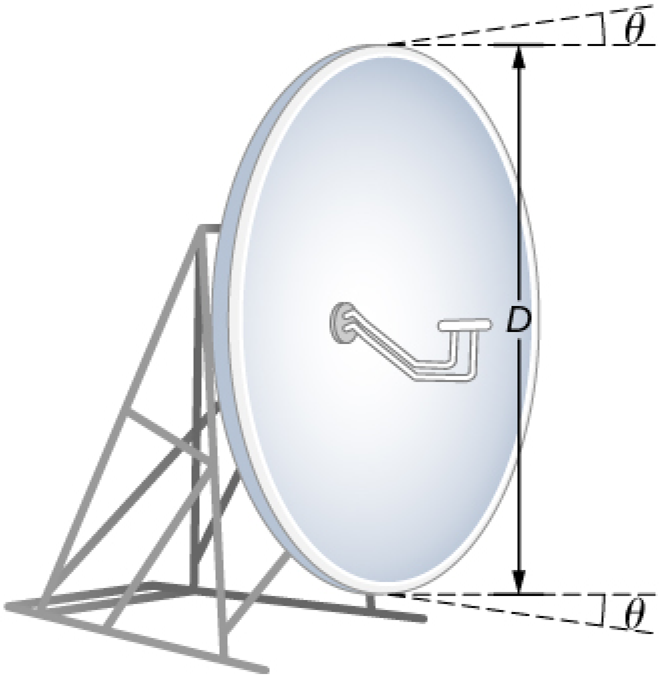
Problem:

Check Your Understanding What is the angular resolution of the Arecibo telescope shown in [link] when operated at 21-cm wavelength? How does it compare to the resolution of the Hubble Telescope?

Solution:

$8.4\ \times\ 10^{-4}$ rad, 3000 times broader than the Hubble Telescope

Diffraction is not only a problem for optical instruments but also for the electromagnetic radiation itself. Any beam of light having a finite diameter *D* and a wavelength λ exhibits diffraction spreading. The beam spreads out with an angle θ given by [link], $\theta = 1.22\lambda/D$. Take, for example, a laser beam made of rays as parallel as possible (angles between rays as close to $\theta = 0^\circ$ as possible) instead spreads out at an angle $\theta = 1.22\lambda/D$, where D is the diameter of the beam and λ is its wavelength. This spreading is impossible to observe for a flashlight because its beam is not very parallel to start with. However, for long-distance transmission of laser beams or microwave signals, diffraction spreading can be significant ([link]). To avoid this, we can increase *D*. This is done for laser light sent to the moon to measure its distance from Earth. The laser beam is expanded through a telescope to make *D* much larger and θ smaller.



The beam produced by this microwave transmission antenna spreads out at a minimum angle $\theta = 1.22\lambda/D$ due to diffraction. It is impossible to produce a near-parallel beam because the beam has a limited diameter.

In most biology laboratories, resolution is an issue when the use of the microscope is introduced. The smaller the distance *x* by which two objects can be separated and still be seen as distinct, the greater the resolution. The resolving power of a lens is defined as that distance *x*. An expression for resolving power is obtained from the Rayleigh criterion. [link](a) shows two point objects separated by a distance *x*. According to the Rayleigh criterion, resolution is possible when the minimum angular separation is

Equation:

$$\theta = 1.22\frac{\lambda}{D} = \frac{x}{d},$$

where *d* is the distance between the specimen and the objective lens, and we have used the small angle approximation (i.e., we have assumed that *x* is much smaller than *d*), so that $\tan\theta \approx \sin\theta \approx \theta$. Therefore, the resolving power is

Equation:

$$x = 1.22\frac{\lambda d}{D}.$$

Another way to look at this is by the concept of numerical aperture (*NA*), which is a measure of the maximum acceptance angle at which a lens will take light and still contain it within the lens. [link](b) shows a lens and an object at point *P*. The *NA* here is a measure of the ability of the lens to gather light and resolve fine detail. The angle subtended by the lens at its focus is defined to be $\theta = 2\alpha$. From the figure and again using the small angle approximation, we can write

Equation:

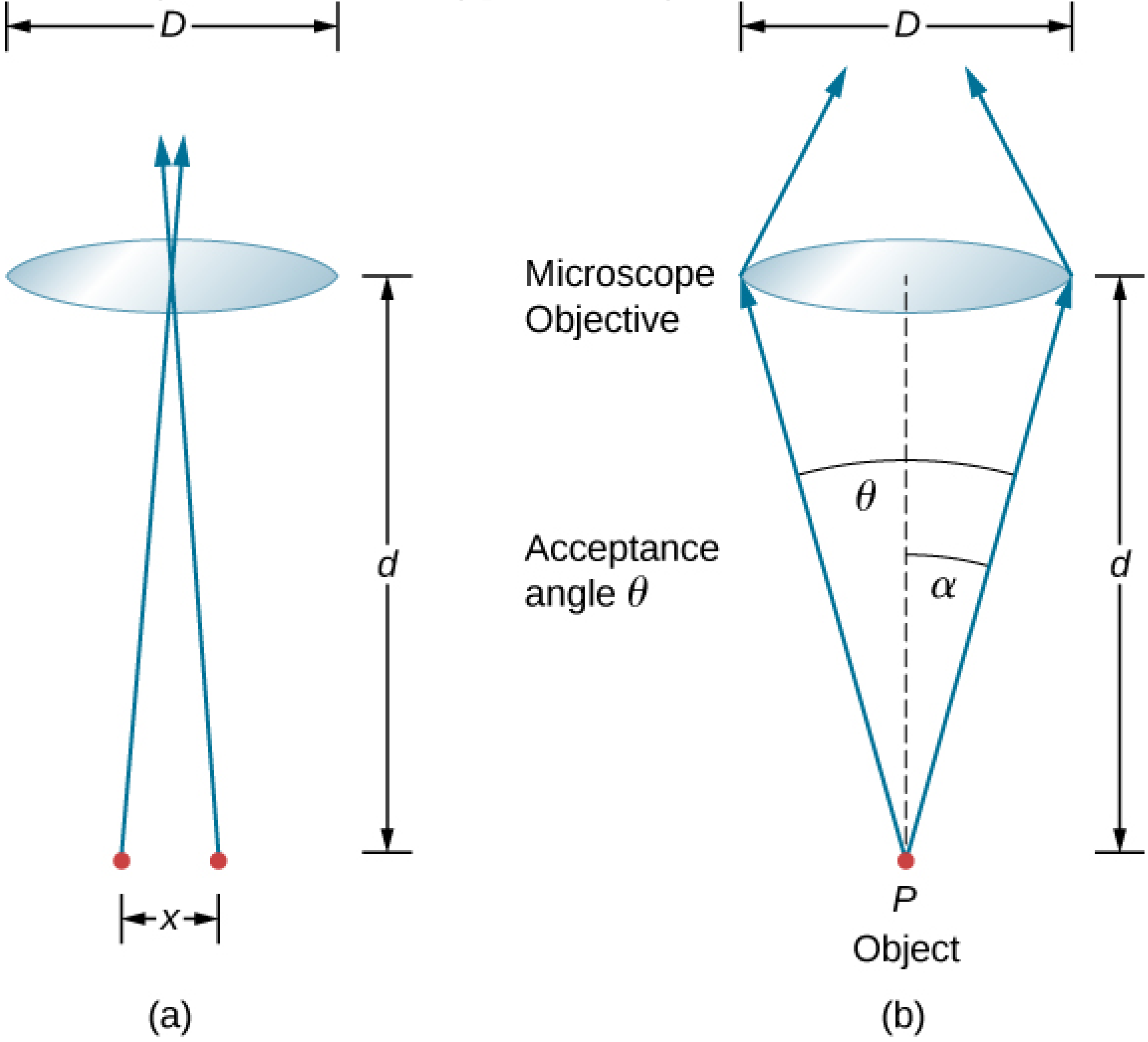
$$\sin\alpha = \frac{D/2}{d} = \frac{D}{2d}.$$

The *NA* for a lens is $NA = n\sin\alpha$, where *n* is the index of refraction of the medium between the objective lens and the object at point *P*. From this definition for *NA*, we can see that

Equation:

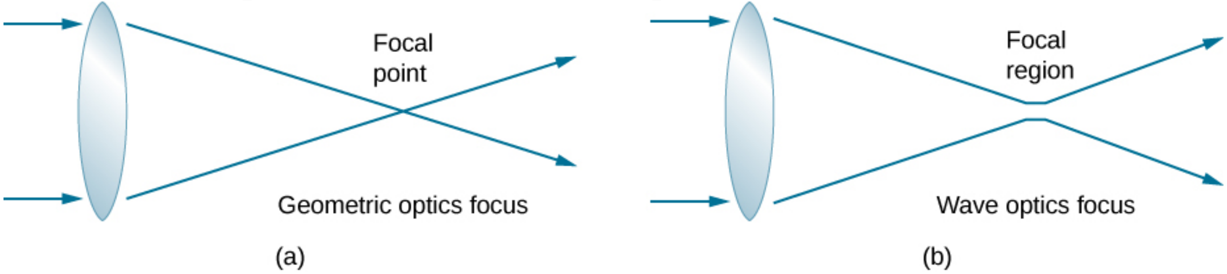
$$x = 1.22\frac{\lambda d}{D} = 1.22\frac{\lambda}{2\sin\alpha} = 0.61\frac{\lambda n}{NA}.$$

In a microscope, *NA* is important because it relates to the resolving power of a lens. A lens with a large *NA* is able to resolve finer details. Lenses with larger *NA* are also able to collect more light and so give a brighter image. Another way to describe this situation is that the larger the *NA*, the larger the cone of light that can be brought into the lens, so more of the diffraction modes are collected. Thus the microscope has more information to form a clear image, and its resolving power is higher.



(a) Two points separated by a distance *x* and positioned a distance *d* away from the objective. (b) Terms and symbols used in discussion of resolving power for a lens and an object at point *P* (credit a: modification of work by “Infopro”/Wikimedia Commons).

One of the consequences of diffraction is that the focal point of a beam has a finite width and intensity distribution. Imagine focusing when only considering geometric optics, as in [link](a). The focal point is regarded as an infinitely small point with a huge intensity and the capacity to incinerate most samples, irrespective of the *NA* of the objective lens—an unphysical oversimplification. For wave optics, due to diffraction, we take into account the phenomenon in which the focal point spreads to become a focal spot ([link](b)) with the size of the spot decreasing with increasing *NA*. Consequently, the intensity in the focal spot increases with increasing *NA*. The higher the *NA*, the greater the chances of photodegrading the specimen. However, the spot never becomes a true point.



(a) In geometric optics, the focus is modelled as a point, but it is not physically possible to produce such a point because it implies infinite intensity. (b) In wave optics, the focus is an extended region.

In a different type of microscope, molecules within a specimen are made to emit light through a mechanism called fluorescence. By controlling the molecules emitting light, it has become possible to construct images with resolution much finer than the Rayleigh criterion, thus circumventing the diffraction limit. The development of super-resolved fluorescence microscopy led to the 2014 Nobel Prize in Chemistry.

Note:
In this Optical Resolution Model, two diffraction patterns for light through two circular apertures are shown side by side in this simulation by Fu-Kwun Hwang. Watch the patterns merge as you decrease the aperture diameters.

Summary

- Diffraction limits resolution.
- The Rayleigh criterion states that two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other.

Conceptual Questions

Exercise:

Problem:

Is higher resolution obtained in a microscope with red or blue light? Explain your answer.

Solution:

blue; The shorter wavelength of blue light results in a smaller angle for diffraction limit.

Exercise:

Problem:

The resolving power of refracting telescope increases with the size of its objective lens. What other advantage is gained with a larger lens?

Exercise:

Problem:

The distance between atoms in a molecule is about 10^{-8} cm. Can visible light be used to “see” molecules?

Solution:

No, these distances are three orders of magnitude smaller than the wavelength of visible light, so visible light makes a poor probe for atoms.

Exercise:

Problem:

A beam of light always spreads out. Why can a beam not be created with parallel rays to prevent spreading? Why can lenses, mirrors, or apertures not be used to correct the spreading?

Problems

Exercise:

Problem:

The 305-m-diameter Arecibo radio telescope pictured in [link] detects radio waves with a 4.00-cm average wavelength. (a) What is the angle between two just-resolvable point sources for this telescope? (b) How close together could these point sources be at the 2 million light-year distance of the Andromeda Galaxy?

Exercise:

Problem:

Assuming the angular resolution found for the Hubble Telescope in [link], what is the smallest detail that could be observed on the moon?

Solution:

107 m

Exercise:

Problem:

Diffraction spreading for a flashlight is insignificant compared with other limitations in its optics, such as spherical aberrations in its mirror. To show this, calculate the minimum angular spreading of a flashlight beam that is originally 5.00 cm in diameter with an average wavelength of 600 nm.

Exercise:

Problem:

(a) What is the minimum angular spread of a 633-nm wavelength He-Ne laser beam that is originally 1.00 mm in diameter? (b) If this laser is aimed at a mountain cliff 15.0 km away, how big will the illuminated spot be? (c) How big a spot would be illuminated on the moon, neglecting atmospheric effects? (This might be done to hit a corner reflector to measure the round-trip time and, hence, distance.)

Solution:

a. $7.72 \times 10^{-4}\,\text{rad}$; b. 23.2 m; c. 590 km

Exercise:

Problem:

A telescope can be used to enlarge the diameter of a laser beam and limit diffraction spreading. The laser beam is sent through the telescope in opposite the normal direction and can then be projected onto a satellite or the moon. (a) If this is done with the Mount Wilson telescope, producing a 2.54-m-diameter beam of 633-nm light, what is the minimum angular spread of the beam? (b) Neglecting atmospheric effects, what is the size of the spot this beam would make on the moon, assuming a lunar distance of $3.84 \times 10^{8}\,\text{m}$?

Exercise:

Problem:

The limit to the eye's acuity is actually related to diffraction by the pupil. (a) What is the angle between two just-resolvable points of light for a 3.00-mm-diameter pupil, assuming an average wavelength of 550 nm? (b) Take your result to be the practical limit for the eye. What is the greatest possible distance a car can be from you if you can resolve its two headlights, given they are 1.30 m apart? (c) What is the distance between two just-resolvable points held at an arm's length (0.800 m) from your eye? (d) How does your answer to (c) compare to details you normally observe in everyday circumstances?

Solution:

a. 2.24×10^{-4} rad; b. 5.81 km; c. 0.179 mm; d. can resolve details 0.2 mm apart at arm's length

Exercise:

Problem:

What is the minimum diameter mirror on a telescope that would allow you to see details as small as 5.00 km on the moon some 384,000 km away? Assume an average wavelength of 550 nm for the light received.

Exercise:

Problem:

Find the radius of a star's image on the retina of an eye if its pupil is open to 0.65 cm and the distance from the pupil to the retina is 2.8 cm. Assume $\lambda = 550\,\text{nm}$.

Solution:

$2.9\ \mu\text{m}$

Exercise:

Problem:

(a) The dwarf planet Pluto and its moon, Charon, are separated by 19,600 km. Neglecting atmospheric effects, should the 5.08-m-diameter Palomar Mountain telescope be able to resolve these bodies when they are 4.50×10^9 km from Earth? Assume an average wavelength of 550 nm. (b) In actuality, it is just barely possible to discern that Pluto and Charon are separate bodies using a ground-based telescope. What are the reasons for this?

Exercise:

Problem:

A spy satellite orbits Earth at a height of 180 km. What is the minimum diameter of the objective lens in a telescope that must be used to resolve columns of troops marching 2.0 m apart? Assume $\lambda = 550\,\text{nm}$.

Solution:

6.0 cm

Exercise:

Problem:

What is the minimum angular separation of two stars that are just-resolvable by the 8.1-m Gemini South telescope, if atmospheric effects do not limit resolution? Use 550 nm for the wavelength of the light from the stars.

Exercise:

Problem:

The headlights of a car are 1.3 m apart. What is the maximum distance at which the eye can resolve these two headlights? Take the pupil diameter to be 0.40 cm.

Solution:

7.71 km

Exercise:

Problem:

When dots are placed on a page from a laser printer, they must be close enough so that you do not see the individual dots of ink. To do this, the separation of the dots must be less than Raleigh's criterion. Take the pupil of the eye to be 3.0 mm and the distance from the paper to the eye of 35 cm; find the minimum separation of two dots such that they cannot be resolved. How many dots per inch (dpi) does this correspond to?

Exercise:

Problem:

Suppose you are looking down at a highway from a jetliner flying at an altitude of 6.0 km. How far apart must two cars be if you are able to distinguish them? Assume that $\lambda = 550\ \text{nm}$ and that the diameter of your pupils is 4.0 mm.

Solution:

1.0 m

Exercise:

Problem:

Can an astronaut orbiting Earth in a satellite at a distance of 180 km from the surface distinguish two skyscrapers that are 20 m apart? Assume that the pupils of the astronaut's eyes have a diameter of 5.0 mm and that most of the light is centered around 500 nm.

Exercise:

Problem:

The characters of a stadium scoreboard are formed with closely spaced lightbulbs that radiate primarily yellow light. (Use $\lambda = 600 \text{ nm}$.) How closely must the bulbs be spaced so that an observer 80 m away sees a display of continuous lines rather than the individual bulbs? Assume that the pupil of the observer's eye has a diameter of 5.0 mm.

Solution:

1.2 cm or closer

Exercise:

Problem:

If a microscope can accept light from objects at angles as large as $\alpha = 70^\circ$, what is the smallest structure that can be resolved when illuminated with light of wavelength 500 nm and (a) the specimen is in air? (b) When the specimen is immersed in oil, with index of refraction of 1.52?

Exercise:

Problem:

A camera uses a lens with aperture 2.0 cm. What is the angular resolution of a photograph taken at 700 nm wavelength? Can it resolve the millimeter markings of a ruler placed 35 m away?

Solution:

no

Glossary

diffraction limit
: fundamental limit to resolution due to diffraction

Rayleigh criterion

two images are just-resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other

resolution

ability, or limit thereof, to distinguish small details in images

X-Ray Diffraction

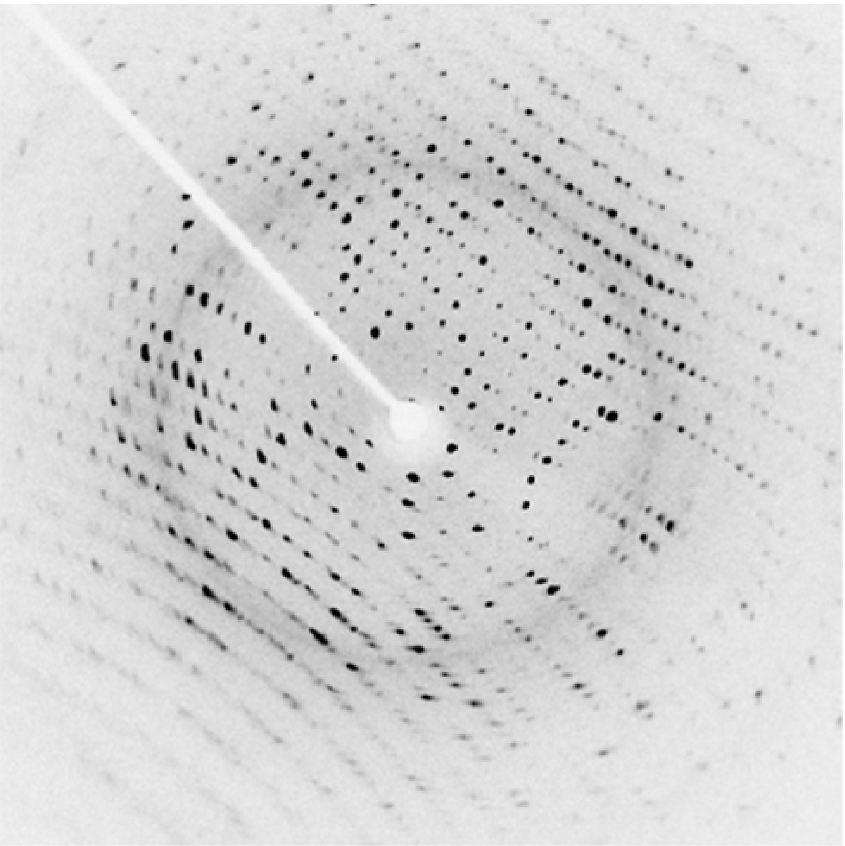
By the end of this section, you will be able to:

- Describe interference and diffraction effects exhibited by X-rays in interaction with atomic-scale structures

Since X-ray photons are very energetic, they have relatively short wavelengths, on the order of 10^{-8} m to 10^{-12} m. Thus, typical X-ray photons act like rays when they encounter macroscopic objects, like teeth, and produce sharp shadows. However, since atoms are on the order of 0.1 nm in size, X-rays can be used to detect the location, shape, and size of atoms and molecules. The process is called **X-ray diffraction**, and it involves the interference of X-rays to produce patterns that can be analyzed for information about the structures that scattered the X-rays.

Perhaps the most famous example of X-ray diffraction is the discovery of the double-helical structure of DNA in 1953 by an international team of scientists working at England’s Cavendish Laboratory—American James Watson, Englishman Francis Crick, and New Zealand-born Maurice Wilkins. Using X-ray diffraction data produced by Rosalind Franklin, they were the first to model the double-helix structure of DNA that is so crucial to life. For this work, Watson, Crick, and Wilkins were awarded the 1962 Nobel Prize in Physiology or Medicine. (There is some debate and controversy over the issue that Rosalind Franklin was not included in the prize, although she died in 1958, before the prize was awarded.)

[link] shows a diffraction pattern produced by the scattering of X-rays from a crystal. This process is known as X-ray crystallography because of the information it can yield about crystal structure, and it was the type of data Rosalind Franklin supplied to Watson and Crick for DNA. Not only do X-rays confirm the size and shape of atoms, they give information about the atomic arrangements in materials. For example, more recent research in high-temperature superconductors involves complex materials whose lattice arrangements are crucial to obtaining a superconducting material. These can be studied using X-ray crystallography.



X-ray diffraction from the crystal of a protein (hen egg lysozyme) produced this interference pattern. Analysis of the pattern yields information about the structure of the protein. (credit: “Del45”/Wikimedia Commons)

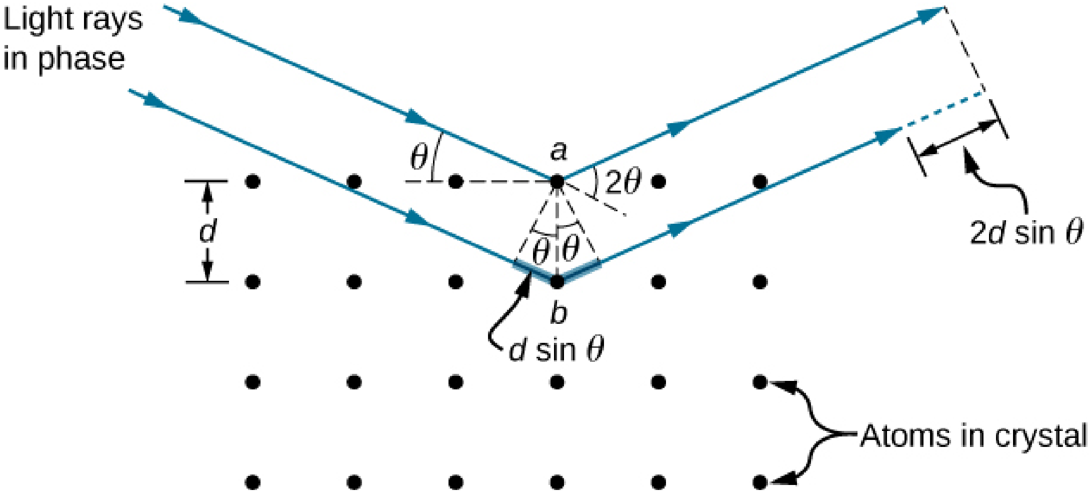
Historically, the scattering of X-rays from crystals was used to prove that X-rays are energetic electromagnetic (EM) waves. This was suspected from the time of the discovery of X-rays in 1895, but it was not until 1912 that the German Max von Laue (1879–1960) convinced two of his colleagues to scatter X-rays from crystals. If a diffraction pattern is obtained, he reasoned, then the X-rays must be waves, and their wavelength could be determined. (The spacing of atoms in various crystals was reasonably well known at the time, based on good values for Avogadro’s number.) The experiments were convincing, and the 1914 Nobel Prize in Physics was given to von Laue for his suggestion leading to the proof that X-rays are EM waves. In 1915, the unique father-and-son team of Sir William Henry Bragg and his son Sir William Lawrence Bragg were awarded a joint Nobel Prize for inventing the X-ray spectrometer and the then-new science of X-ray analysis.

In ways reminiscent of thin-film interference, we consider two plane waves at X-ray wavelengths, each one reflecting off a different plane of atoms within a crystal's lattice, as shown in [link]. From the geometry, the difference in path lengths is $2d\sin\theta$. Constructive interference results when this distance is an integer multiple of the wavelength. This condition is captured by the *Bragg equation*,

Note:
Equation:

$$m\lambda = 2d\sin\theta,\ m = 1,\ 2,\ 3\ldots$$

where *m* is a positive integer and *d* is the spacing between the planes. Following the Law of Reflection, both the incident and reflected waves are described by the same angle, θ, but unlike the general practice in geometric optics, θ is measured with respect to the surface itself, rather than the normal.



X-ray diffraction with a crystal. Two incident waves reflect off two planes of a crystal. The difference in path lengths is indicated by the dashed line.

Example:
X-Ray Diffraction with Salt Crystals

Common table salt is composed mainly of NaCl crystals. In a NaCl crystal, there is a family of planes 0.252 nm apart. If the first-order maximum is observed at an incidence angle of 18.1°, what is the wavelength of the X-ray scattering from this crystal?

Strategy

Use the Bragg equation, [link], $m\lambda = 2d\sin\theta$, to solve for θ.

Solution

For first-order, $m = 1$, and the plane spacing *d* is known. Solving the Bragg equation for wavelength yields

Equation:

$$\lambda = \frac{2d\sin\theta}{m} = \frac{2\left(0.252 \times 10^{-9}\,\text{m}\right)\sin\left(18.1^\circ\right)}{1} = 1.57 \times 10^{-10}\,\text{m},\ \text{or}\ 0.157\,\text{nm}.$$

Significance

The determined wavelength fits within the X-ray region of the electromagnetic spectrum. Once again, the wave nature of light makes itself prominent when the wavelength $(\lambda = 0.157\,\text{nm})$ is comparable to the size of the physical structures $(d = 0.252\,\text{nm})$ it interacts with.

Note:

Exercise:

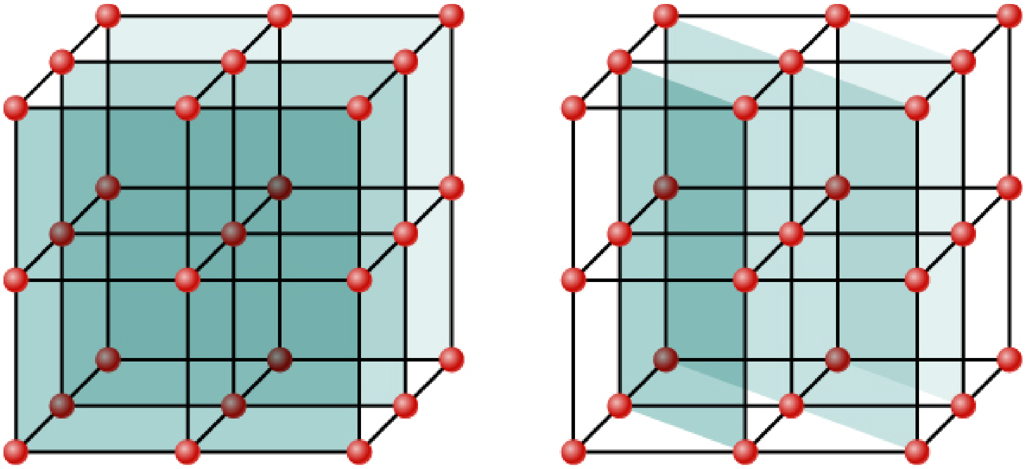
Problem:

Check Your Understanding For the experiment described in [link], what are the two other angles where interference maxima may be observed? What limits the number of maxima?

Solution:

38.4° and 68.8°; Between $\theta = 0^\circ \rightarrow 90^\circ$, orders 1, 2, and 3, are all that exist.

Although [link] depicts a crystal as a two-dimensional array of scattering centers for simplicity, real crystals are structures in three dimensions. Scattering can occur simultaneously from different families of planes at different orientations and spacing patterns known as called **Bragg planes**, as shown in [link]. The resulting interference pattern can be quite complex.



Because of the regularity that makes a crystal structure, one crystal can have many families of planes within its geometry, each one giving rise to X-ray diffraction.

Summary

- X-rays are relatively short-wavelength EM radiation and can exhibit wave characteristics such as interference when interacting with correspondingly small objects.

Conceptual Questions

Exercise:

Problem: Crystal lattices can be examined with X-rays but not UV. Why?

Solution:

UV wavelengths are much larger than lattice spacings in crystals such that there is no diffraction. The Bragg equation implies a value for $sin\theta$ greater than unity, which has no solution.

Problems

Exercise:

Problem:

X-rays of wavelength 0.103 nm reflects off a crystal and a second-order maximum is recorded at a Bragg angle of 25.5°. What is the spacing between the scattering planes in this crystal?

Exercise:

Problem:

A first-order Bragg reflection maximum is observed when a monochromatic X-ray falls on a crystal at a 32.3° angle to a reflecting plane. What is the wavelength of this X-ray?

Solution:

0.120 nm

Exercise:

Problem:

An X-ray scattering experiment is performed on a crystal whose atoms form planes separated by 0.440 nm. Using an X-ray source of wavelength 0.548 nm, what is the angle (with respect to the planes in question) at which the experimenter needs to illuminate the crystal in order to observe a first-order maximum?

Exercise:

Problem:

The structure of the NaCl crystal forms reflecting planes 0.541 nm apart. What is the smallest angle, measured from these planes, at which X-ray diffraction can be observed, if X-rays of wavelength 0.085 nm are used?

Solution:

4.51°

Exercise:

Problem:

On a certain crystal, a first-order X-ray diffraction maximum is observed at an angle of 27.1° relative to its surface, using an X-ray source of unknown wavelength. Additionally, when illuminated with a different, this time of known wavelength 0.137 nm, a second-order maximum is detected at 37.3°. Determine (a) the spacing between the reflecting planes, and (b) the unknown wavelength.

Exercise:

Problem:

Calcite crystals contain scattering planes separated by 0.30 nm. What is the angular separation between first and second-order diffraction maxima when X-rays of 0.130 nm wavelength are used?

Solution:

13.2°

Exercise:

Problem:

The first-order Bragg angle for a certain crystal is 12.1°. What is the second-order angle?

Glossary

Bragg planes
families of planes within crystals that can give rise to X-ray diffraction

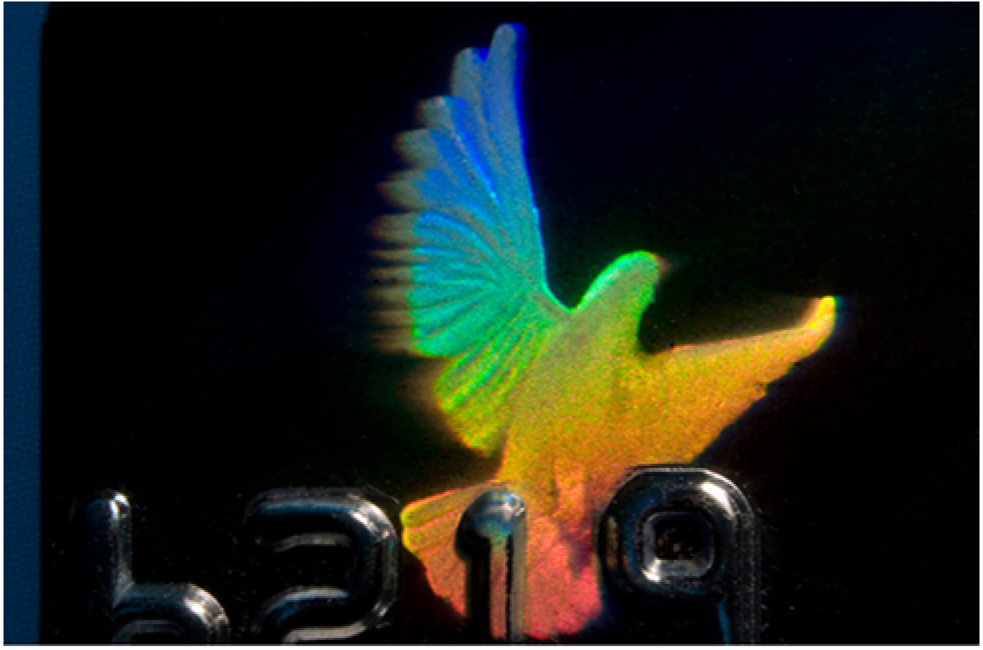
X-ray diffraction
technique that provides the detailed information about crystallographic structure of natural and manufactured materials

Holography

By the end of this section, you will be able to:

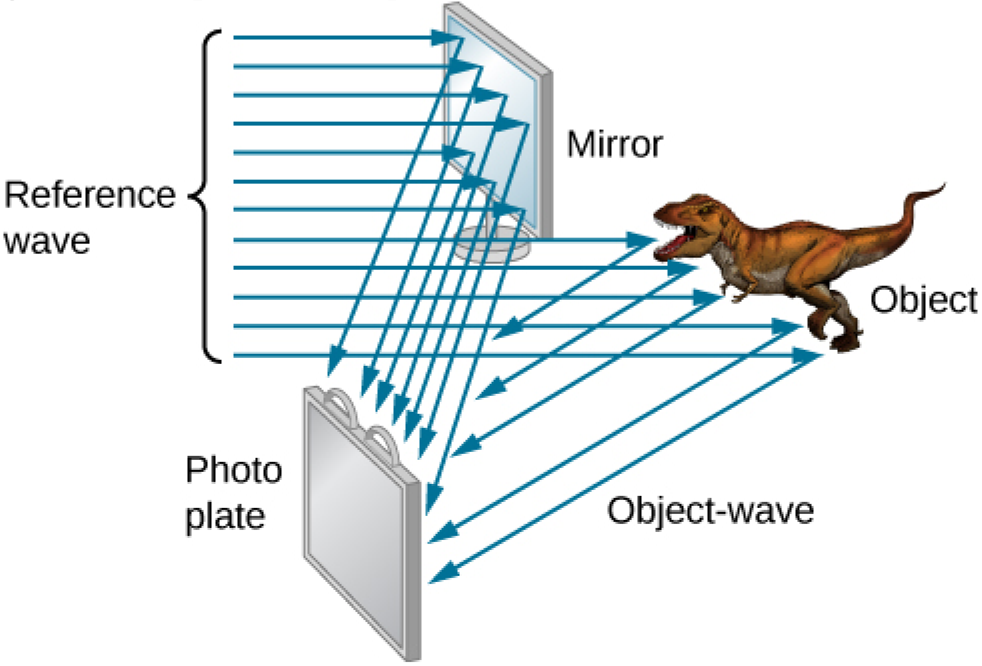
- Describe how a three-dimensional image is recorded as a hologram
- Describe how a three-dimensional image is formed from a hologram

A **hologram**, such as the one in [link], is a true three-dimensional image recorded on film by lasers. Holograms are used for amusement; decoration on novelty items and magazine covers; security on credit cards and driver's licenses (a laser and other equipment are needed to reproduce them); and for serious three-dimensional information storage. You can see that a hologram is a true three-dimensional image because objects change relative position in the image when viewed from different angles.



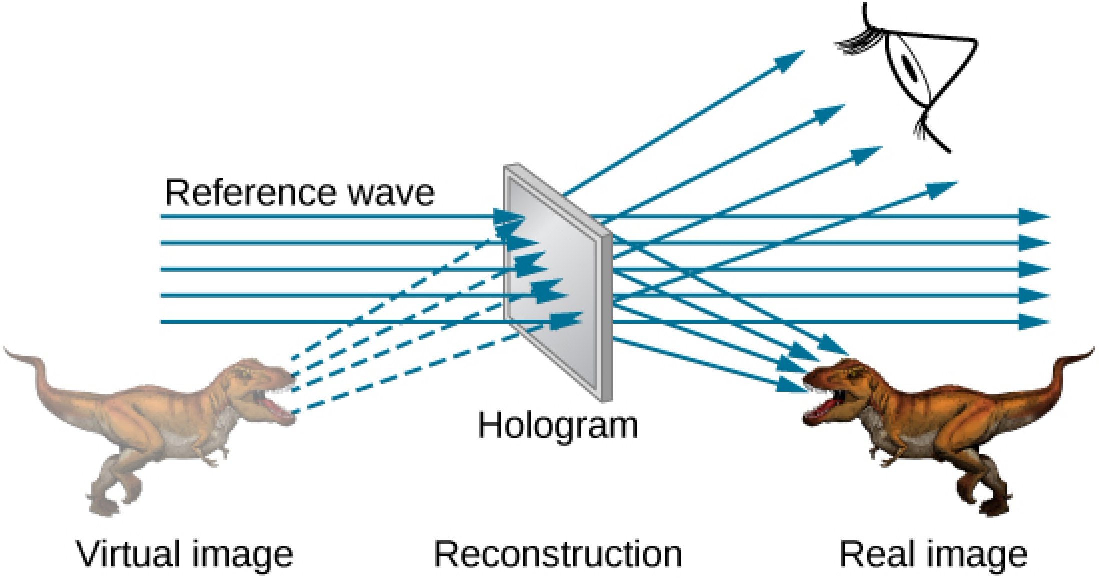
Credit cards commonly have holograms for logos, making them difficult to reproduce. (credit: Dominic Alves)

The name hologram means "entire picture" (from the Greek *holo*, as in holistic) because the image is three-dimensional. **Holography** is the process of producing holograms and, although they are recorded on photographic film, the process is quite different from normal photography. Holography uses light interference or wave optics, whereas normal photography uses geometric optics. [link] shows one method of producing a hologram. Coherent light from a laser is split by a mirror, with part of the light illuminating the object. The remainder, called the reference beam, shines directly on a piece of film. Light scattered from the object interferes with the reference beam, producing constructive and destructive interference. As a result, the exposed film looks foggy, but close examination reveals a complicated interference pattern stored on it. Where the interference was constructive, the film (a negative actually) is darkened. Holography is sometimes called lens-less photography, because it uses the wave characteristics of light, as contrasted to normal photography, which uses geometric optics and requires lenses.



Production of a hologram. Single-wavelength coherent light from a laser produces a well-defined interference pattern on a piece of film. The laser beam is split by a partially silvered mirror, with part of the light illuminating the object and the remainder shining directly on the film. (credit: modification of work by Mariana Ruiz Villarreal)

Light falling on a hologram can form a three-dimensional image of the original object. The process is complicated in detail, but the basics can be understood, as shown in [link], in which a laser of the same type that exposed the film is now used to illuminate it. The myriad tiny exposed regions of the film are dark and block the light, whereas less exposed regions allow light to pass. The film thus acts much like a collection of diffraction gratings with various spacing patterns. Light passing through the hologram is diffracted in various directions, producing both real and virtual images of the object used to expose the film. The interference pattern is the same as that produced by the object. Moving your eye to various places in the interference pattern gives you different perspectives, just as looking directly at the object would. The image thus looks like the object and is three dimensional like the object.



A transmission hologram is one that produces real and virtual images when a laser of the same type as that which exposed the hologram is passed through it. Diffraction from various parts of the film produces the same interference pattern that was produced by the object that was used to expose it. (credit: modification of work by Mariana Ruiz Villarreal)

The hologram illustrated in [link] is a transmission hologram. Holograms that are viewed with reflected light, such as the white light holograms on credit cards, are reflection holograms and are more common. White light holograms often appear a little blurry with rainbow edges, because the diffraction patterns of various colors of light are at slightly different locations due to their different wavelengths. Further uses of holography include all types of three-dimensional information storage, such as of statues in museums, engineering studies of structures, and images of human organs.

Invented in the late 1940s by Dennis Gabor (1900–1970), who won the 1971 Nobel Prize in Physics for his work, holography became far more practical with the development of the laser. Since lasers produce coherent single-wavelength light, their interference patterns are more pronounced. The precision is so great that it is even possible to record numerous holograms on a single piece of film by just changing the angle of the film for each successive image. This is how the holograms that move as you walk by them are produced—a kind of lens-less movie.

In a similar way, in the medical field, holograms have allowed complete three-dimensional holographic displays of objects from a stack of images. Storing these images for future use is relatively easy. With the use of an endoscope, high-resolution, three-dimensional holographic images of internal organs and tissues can be made.

Summary

- Holography is a technique based on wave interference to record and form three-dimensional images.
- Lasers offer a practical way to produce sharp holographic images because of their monochromatic and coherent light for pronounced interference patterns.

Key Equations

Destructive interference for a single slit	$D\sin\theta = m\lambda\,\text{for}\,m = \pm1,\ \pm2,\ \pm3, \ldots$
Half phase angle	$\beta = \frac{\phi}{2} = \frac{\pi D\sin\theta}{\lambda}$
Field amplitude in the diffraction pattern	$E = N\Delta E_0\frac{\sin\beta}{\beta}$
Intensity in the diffraction pattern	$I = I_0\left(\frac{\sin\beta}{\beta}\right)^2$
Rayleigh criterion for circular apertures	$\theta = 1.22\frac{\lambda}{D}$
Bragg equation	$m\lambda = 2d\sin\theta,\ m = 1,\ 2,\ 3\ldots$

Conceptual Questions

Exercise:

Problem:

How can you tell that a hologram is a true three-dimensional image and that those in three-dimensional movies are not?

Exercise:

Problem:

If a hologram is recorded using monochromatic light at one wavelength but its image is viewed at another wavelength, say 10% shorter, what will you see? What if it is viewed using light of exactly half the original wavelength?

Solution:

Image will appear at slightly different location and/or size when viewed using 10% shorter wavelength but at exactly half the wavelength, a higher-order interference reconstructs the original image, different color.

Exercise:

Problem:

What image will one see if a hologram is recorded using monochromatic light but its image is viewed in white light? Explain.

Additional Problems

Exercise:

Problem:

White light falls on two narrow slits separated by 0.40 mm. The interference pattern is observed on a screen 3.0 m away. (a) What is the separation between the first maxima for red light $(\lambda = 700\,\text{nm})$ and violet light $(\lambda = 400\,\text{nm})$? (b) At what point nearest the central maximum will a maximum for yellow light $(\lambda = 600\,\text{nm})$ coincide with a maximum for violet light? Identify the order for each maximum.

Solution:

a. 2.2 mm; b. 0.172°, second-order yellow and third-order violet coincide

Exercise:

Problem:

Microwaves of wavelength 10.0 mm fall normally on a metal plate that contains a slit 25 mm wide. (a) Where are the first minima of the diffraction pattern? (b) Would there be minima if the wavelength were 30.0 mm?

Exercise:

Problem:

Quasars, or *quasi-stellar radio sources*, are astronomical objects discovered in 1960. They are distant but strong emitters of radio waves with angular size so small, they were originally unresolved, the same as stars. The quasar 3C405 is actually two discrete radio sources that subtend an angle of 82 arcsec. If this object is studied using radio emissions at a frequency of 410 MHz, what is the minimum diameter of a radio telescope that can resolve the two sources?

Solution:

2.2 km

Exercise:

Problem:

Two slits each of width 1800 nm and separated by the center-to-center distance of 1200 nm are illuminated by plane waves from a krypton ion laser-emitting at wavelength 461.9 nm. Find the number of interference peaks in the central diffraction peak.

Exercise:

Problem:

A microwave of an unknown wavelength is incident on a single slit of width 6 cm. The angular width of the central peak is found to be 25°. Find the wavelength.

Solution:

1.3 cm

Exercise:

Problem:

Red light (wavelength 632.8 nm in air) from a Helium-Neon laser is incident on a single slit of width 0.05 mm. The entire apparatus is immersed in water of refractive index 1.333. Determine the angular width of the central peak.

Exercise:

Problem:

A light ray of wavelength 461.9 nm emerges from a 2-mm circular aperture of a krypton ion laser. Due to diffraction, the beam expands as it moves out. How large is the central bright spot at (a) 1 m, (b) 1 km, (c) 1000 km, and (d) at the surface of the moon at a distance of 400,000 km from Earth.

Solution:

a. 0.28 mm; b. 0.28 m; c. 280 m; d. 113 km

Exercise:

Problem:

How far apart must two objects be on the moon to be distinguishable by eye if only the diffraction effects of the eye's pupil limit the resolution? Assume 550 nm for the wavelength of light, the pupil diameter 5.0 mm, and 400,000 km for the distance to the moon.

Exercise:

Problem:

How far apart must two objects be on the moon to be resolvable by the 8.1-m-diameter Gemini North telescope at Mauna Kea, Hawaii, if only the diffraction effects of the telescope aperture limit the resolution? Assume 550 nm for the wavelength of light and 400,000 km for the distance to the moon.

Solution:

33 m

Exercise:

Problem:

A spy satellite is reputed to be able to resolve objects 10. cm apart while operating 197 km above the surface of Earth. What is the diameter of the aperture of the telescope if the resolution is only limited by the diffraction effects? Use 550 nm for light.

Exercise:

Problem:

Monochromatic light of wavelength 530 nm passes through a horizontal single slit of width $1.5\ \mu\text{m}$ in an opaque plate. A screen of dimensions $2.0\ \text{m}\ \times\ 2.0\ \text{m}$ is 1.2 m away from the slit. (a) Which way is the diffraction pattern spread out on the screen? (b) What are the angles of the minima with respect to the center? (c) What are the angles of the maxima? (d) How wide is the central bright fringe on the screen? (e) How wide is the next bright fringe on the screen?

Solution:

a. vertically; b. $\pm 20^\circ$, $\pm 44^\circ$; c. 0, $\pm 31^\circ$, $\pm 60^\circ$; d. 89 cm; e. 71 cm

Exercise:

Problem:

A monochromatic light of unknown wavelength is incident on a slit of width $20\ \mu\text{m}$. A diffraction pattern is seen at a screen 2.5 m away where the central maximum is spread over a distance of 10.0 cm. Find the wavelength.

Exercise:

Problem:

A source of light having two wavelengths 550 nm and 600 nm of equal intensity is incident on a slit of width $1.8\ \mu\text{m}$. Find the separation of the $m = 1$ bright spots of the two wavelengths on a screen 30.0 cm away.

Solution:

0.98 cm

Exercise:

Problem:

A single slit of width 2100 nm is illuminated normally by a wave of wavelength 632.8 nm. Find the phase difference between waves from the top and one third from the bottom of the slit to a point on a screen at a horizontal distance of 2.0 m and vertical distance of 10.0 cm from the center.

Exercise:

Problem:

A single slit of width $3.0\ \mu\text{m}$ is illuminated by a sodium yellow light of wavelength 589 nm. Find the intensity at a 15° angle to the axis in terms of the intensity of the central maximum.

Solution:

$I/I_0 = 0.041$

Exercise:

Problem:

A single slit of width 0.10 mm is illuminated by a mercury lamp of wavelength 576 nm. Find the intensity at a 10° angle to the axis in terms of the intensity of the central maximum.

Exercise:

Problem:

A diffraction grating produces a second maximum that is 89.7 cm from the central maximum on a screen 2.0 m away. If the grating has 600 lines per centimeter, what is the wavelength of the light that produces the diffraction pattern?

Solution:

340 nm

Exercise:

Problem:

A grating with 4000 lines per centimeter is used to diffract light that contains all wavelengths between 400 and 650 nm. How wide is the first-order spectrum on a screen 3.0 m from the grating?

Exercise:

Problem:

A diffraction grating with 2000 lines per centimeter is used to measure the wavelengths emitted by a hydrogen gas discharge tube. (a) At what angles will you find the maxima of the two first-order blue lines of wavelengths 410 and 434 nm? (b) The maxima of two other first-order lines are found at $\theta_1 = 0.097\,\text{rad}$ and $\theta_2 = 0.132\,\text{rad}$. What are the wavelengths of these lines?

Solution:

a. 0.082 rad and 0.087 rad; b. 480 nm and 660 nm

Exercise:

Problem:

For white light $(400\text{ nm} < \lambda < 700\text{ nm})$ falling normally on a diffraction grating, show that the second and third-order spectra overlap no matter what the grating constant *d* is.

Exercise:

Problem:

How many complete orders of the visible spectrum $(400\text{ nm} < \lambda < 700\text{ nm})$ can be produced with a diffraction grating that contains 5000 lines per centimeter?

Solution:

two orders

Exercise:

Problem:

Two lamps producing light of wavelength 589 nm are fixed 1.0 m apart on a wooden plank. What is the maximum distance an observer can be and still resolve the lamps as two separate sources of light, if the resolution is affected solely by the diffraction of light entering the eye? Assume light enters the eye through a pupil of diameter 4.5 mm.

Exercise:

Problem:

On a bright clear day, you are at the top of a mountain and looking at a city 12 km away. There are two tall towers 20.0 m apart in the city. Can your eye resolve the two towers if the diameter of the pupil is 4.0 mm? If not, what should be the minimum magnification power of the telescope needed to resolve the two towers? In your calculations use 550 nm for the wavelength of the light.

Solution:

yes and N/A

Exercise:

Problem:

Radio telescopes are telescopes used for the detection of radio emission from space. Because radio waves have much longer wavelengths than visible light, the diameter of a radio telescope must be very large to provide good resolution. For example, the radio telescope in Penticton, BC in Canada, has a diameter of 26 m and can be operated at frequencies as high as 6.6 GHz. (a) What is the wavelength corresponding to this frequency? (b) What is the angular separation of two radio sources that can be resolved by this telescope? (c) Compare the telescope’s resolution with the angular size of the moon.

(credit: modification of work by Jason Nishiyama)

Exercise:

Problem:

Calculate the wavelength of light that produces its first minimum at an angle of 36.9° when falling on a single slit of width $1.00\ \mu\text{m}$.

Solution:

600 nm

Exercise:

Problem:

(a) Find the angle of the third diffraction minimum for 633-nm light falling on a slit of width $20.0\ \mu\text{m}$. (b) What slit width would place this minimum at 85.0° ?

Exercise:

Problem:

As an example of diffraction by apertures of everyday dimensions, consider a doorway of width 1.0 m. (a) What is the angular position of the first minimum in the diffraction pattern of 600-nm light? (b) Repeat this calculation for a musical note of frequency 440 Hz (A above middle C). Take the speed of sound to be 343 m/s.

Solution:

a. $3.4\ \times\ 10^{-5\,\circ}$; b. 51°

Exercise:

Problem:

What are the angular positions of the first and second minima in a diffraction pattern produced by a slit of width 0.20 mm that is illuminated by 400 nm light? What is the angular width of the central peak?

Exercise:

Problem:

How far would you place a screen from the slit of the previous problem so that the second minimum is a distance of 2.5 mm from the center of the diffraction pattern?

Solution:

0.63 m

Exercise:

Problem:

How narrow is a slit that produces a diffraction pattern on a screen 1.8 m away whose central peak is 1.0 m wide? Assume $\lambda = 589\,\text{nm}$.

Exercise:

Problem:

Suppose that the central peak of a single-slit diffraction pattern is so wide that the first minima can be assumed to occur at angular positions of $\pm 90^\circ$. For this case, what is the ratio of the slit width to the wavelength of the light?

Solution:

1

Exercise:

Problem:

The central diffraction peak of the double-slit interference pattern contains exactly nine fringes. What is the ratio of the slit separation to the slit width?

Exercise:

Problem:

Determine the intensities of three interference peaks other than the central peak in the central maximum of the diffraction, if possible, when a light of wavelength 500 nm is incident normally on a double slit of width 1000 nm and separation 1500 nm. Use the intensity of the central spot to be $1\,\text{mW/cm}^2$.

Solution:

$0.17\ \mathrm{mW/cm^2}$ for $m = 1$ only, no higher orders

Exercise:

Problem:

The yellow light from a sodium vapor lamp *seems* to be of pure wavelength, but it produces two first-order maxima at 36.093° and 36.129° when projected on a 10,000 line per centimeter diffraction grating. What are the two wavelengths to an accuracy of 0.1 nm?

Exercise:

Problem:

Structures on a bird feather act like a reflection grating having 8000 lines per centimeter. What is the angle of the first-order maximum for 600-nm light?

Solution:

28.7°

Exercise:

Problem:

If a diffraction grating produces a first-order maximum for the shortest wavelength of visible light at 30.0°, at what angle will the first-order maximum be for the largest wavelength of visible light?

Exercise:

Problem:

(a) What visible wavelength has its fourth-order maximum at an angle of 25.0° when projected on a 25,000-line per centimeter diffraction grating? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. 42.3 nm; b. This wavelength is not in the visible spectrum. c. The number of slits in this diffraction grating is too large. Etching in integrated circuits can be done to a resolution of 50 nm, so slit separations of 400 nm are at the limit of what we can do today. This line spacing is too small to produce diffraction of light.

Exercise:

Problem:

Consider a spectrometer based on a diffraction grating. Construct a problem in which you calculate the distance between two wavelengths of electromagnetic radiation in your spectrometer. Among the things to be considered are the wavelengths you wish to be able to distinguish, the number of lines per meter on the diffraction grating, and the distance from the grating to the screen or detector. Discuss the practicality of the device in terms of being able to discern between wavelengths of interest.

Exercise:

Problem:

An amateur astronomer wants to build a telescope with a diffraction limit that will allow him to see if there are people on the moons of Jupiter. (a) What diameter mirror is needed to be able to see 1.00-m detail on a Jovian moon at a distance of $7.50 \times 10^{8}\,\text{km}$ from Earth? The wavelength of light averages 600 nm. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. 549 km; b. This is an unreasonably large telescope. c. Unreasonable to assume diffraction limit for optical telescopes unless in space due to atmospheric effects.

Challenge Problems

Exercise:

Problem:

Blue light of wavelength 450 nm falls on a slit of width 0.25 mm. A converging lens of focal length 20 cm is placed behind the slit and focuses the diffraction pattern on a screen. (a) How far is the screen from the lens? (b) What is the distance between the first and the third minima of the diffraction pattern?

Exercise:

Problem:

(a) Assume that the maxima are halfway between the minima of a single-slit diffraction pattern. The use the diameter and circumference of the phasor diagram, as described in Intensity in Single-Slit Diffraction, to determine the intensities of the third and fourth maxima in terms of the intensity of the central maximum. (b) Do the same calculation, using [link].

Solution:

a. $\mathrm{I} = 0.00500\,\mathrm{I}_0,\ 0.00335\,\mathrm{I}_0$; b. $\mathrm{I} = 0.00500\,\mathrm{I}_0,\ 0.00335\,\mathrm{I}_0$

Exercise:

Problem:

(a) By differentiating [link], show that the higher-order maxima of the single-slit diffraction pattern occur at values of β that satisfy $\tan\beta = \beta$. (b) Plot $y = \tan\beta$ and $y = \beta$ versus β and find the intersections of these two curves. What information do they give you about the locations of the maxima? (c) Convince yourself that these points do not appear exactly at $\beta = \left(n + \frac{1}{2}\right)\pi$, where $n = 0,\ 1,\ 2,\ \ldots$, but are quite close to these values.

Exercise:

Problem:

What is the maximum number of lines per centimeter a diffraction grating can have and produce a complete first-order spectrum for visible light?

Solution:

12,800

Exercise:

Problem:

Show that a diffraction grating cannot produce a second-order maximum for a given wavelength of light unless the first-order maximum is at an angle less than 30.0°.

Exercise:

Problem:

A He-Ne laser beam is reflected from the surface of a CD onto a wall. The brightest spot is the reflected beam at an angle equal to the angle of incidence. However, fringes are also observed. If the wall is 1.50 m from the CD, and the first fringe is 0.600 m from the central maximum, what is the spacing of grooves on the CD?

Solution:

$1.58\ \times\ 10^{-6}\,\text{m}$

Exercise:

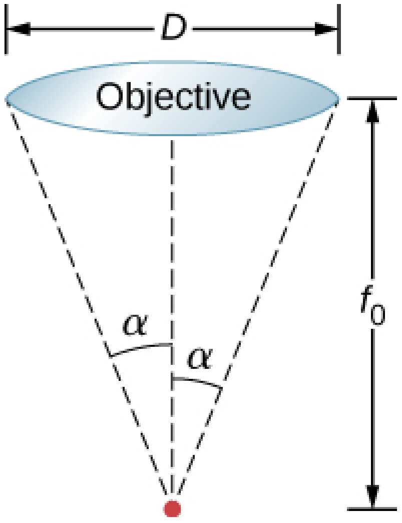
Problem:

Objects viewed through a microscope are placed very close to the focal point of the objective lens. Show that the minimum separation *x* of two objects resolvable through the microscope is given by

Equation:

$$x = \frac{1.22\lambda f_0}{D},$$

where f_0 is the focal length and D is the diameter of the objective lens as shown below.



Glossary

hologram
: three-dimensional image recorded on film by lasers; the word hologram means *entire picture* (from the Greek word *holo*, as in holistic)

holography
: process of producing holograms with the use of lasers

Phasors: Phasor Representation of Signals

Note: This module is part of the collection, *A First Course in Electrical and Computer Engineering*. The LaTeX source files for this collection were created using an optical character recognition technology, and because of this process there may be more errors than usual. Please contact us if you discover any errors.

There are two key ideas behind the phasor representation of a signal:

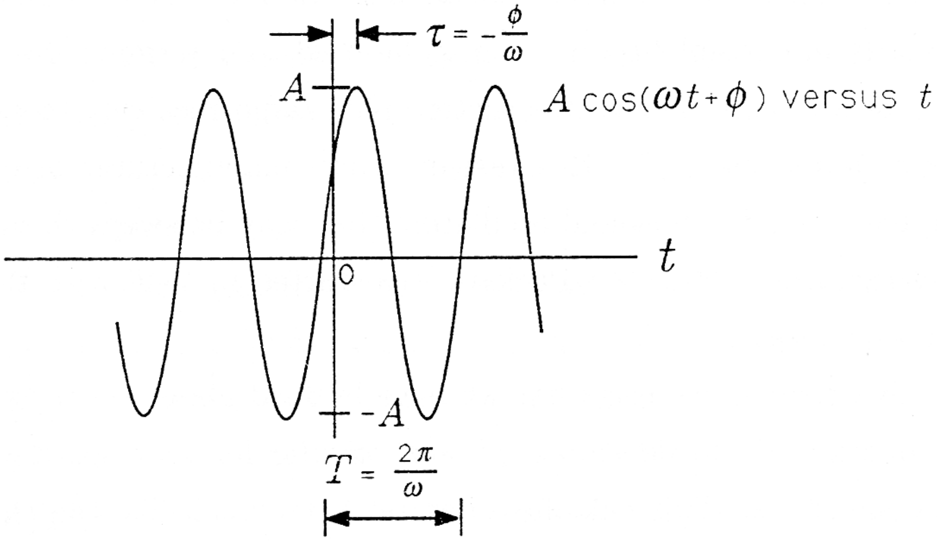
i. a real, time-varying signal may be represented by a complex, time-varying signal; and
ii. a complex, time-varying signal may be represented as the product of a complex number that is *independent* of time and a complex signal that is *dependent* on time.

Let's be concrete. The signal
Equation:

$$x(t) = A \cos\, (\omega t + \varphi),$$

illustrated in [link], is a cosinusoidal signal with amplitude A, frequency ω, and phase φ. The amplitude A characterizes the peak-to-peak swing of $2A$, the angular frequency ω characterizes the period $T = \frac{2\pi}{\omega}$ between negative- to-positive zero crossings (or positive peaks or negative peaks), and the phase φ characterizes the time $\tau = \frac{-\varphi}{\omega}$ when the signal reaches its first peak. With τ so defined, the signal $x(t)$ may also be written as
Equation:

$$x(t) = A \cos\, \omega(t - \tau).$$


$\tau = -\frac{\phi}{\omega}$
A
$A\cos(\omega t+\phi)$ versus t
t
0
$-A$
$T = \frac{2\pi}{\omega}$

A Cosinusoidal Signal

When τ is positive, then τ is a “time delay” that describes the time (greater than zero) when the first peak is achieved. When τ is negative, then τ is a “time advance” that describes the time (less than zero) when the last peak was achieved. With the substitution $\omega = \frac{2\pi}{T}$ we obtain a third way of writing $x(t)$:

Equation:

$$x\left(t\right) = A\cos\frac{2\pi}{T}\left(t-\tau\right).$$

In this form the signal is easy to plot. Simply draw a cosinusoidal wave with amplitude *A* and period *T*; then strike the origin $(t = 0)$ so that the signal reaches its peak at τ. In summary, the parameters that determine a cosinusoidal signal have the following units:

A, arbitrary (e.g., volts or meters/sec, depending upon the application)

ω, in radians/sec (rad/sec)

T, in seconds (sec)

φ, in radians (rad)

τ, in seconds (sec)

Exercise:

Problem:

Show that $x\left(t\right)=A\text{ cos }\frac{2\pi }{T}\left(t-\tau \right)$ is “periodic with period T," meaning that $x(t+mT)=x(t)$ for all integer m.

Exercise:

Problem:

The inverse of the period T is called the “temporal frequency” of the cosinusoidal signal and is given the symbol f; the units of $f=\frac{1}{T}$ are $\left(seconds\right)^{-1}$ or hertz (Hz). Write $x(t)$ in terms of f. How is f related to ω? Explain why f gives the number of cycles of $x(t)$ per second.

Exercise:

Problem:

Sketch the function $x\left(t\right)=110\text{ cos }\left[2\pi \left(60\right)t-\frac{\pi }{8}\right]$ versus t. Repeat for $x\left(t\right)=5\text{ cos }\left[2\pi \left(16\times {10}^{6}\right)t+\frac{\pi }{4}\right]$ and $x\left(t\right)=2\text{ cos }\left[\frac{2\pi }{{10}^{-3}}\left(t-\frac{{10}^{-3}}{8}\right)\right]$. For each function, determine A,ω ,T,f,φ, and τ. Label your sketches carefully.

The signal $x(t)=A\text{ cos }\left(\omega t+\varphi \right)$ can be *represented* as the real part of a complex number:

Equation:

$$\begin{aligned} x\left(t\right) &= \mathrm{Re}\left[Ae^{j\left(\omega t+\varphi\right)}\right] \\ &= \mathrm{Re}\big[Ae^{j\varphi}e^{j\omega t}\big]. \end{aligned}$$

We call $Ae^{j\varphi}e^{j\omega t}$ the complex representation of $x(t)$ and write

Equation:

$$x\left(t\right)\leftrightarrow Ae^{j\varphi}e^{j\omega t},$$

meaning that the signal $x(t)$ may be reconstructed by taking the real part of $Ae^{j\varphi}e^{j\omega t}$. In this representation, we call $Ae^{j\varphi}$ the *phasor* or complex amplitude representation of $x(t)$ and write

Equation:

$$x\left(t\right)\leftrightarrow Ae^{j\varphi},$$

meaning that the signal $x(t)$ may be reconstructed from $Ae^{j\varphi}$ by multiplying with $e^{j\omega t}$ and taking the real part. In communication theory, we call $Ae^{j\varphi}$ the *baseband* representation of the signal $x(t)$.

Exercise:

Problem:

For each of the signals in Problem 3.3, give the corresponding phasor representation $Ae^{j\varphi}$.

Geometric Interpretation. Let's call

Equation:

$$Ae^{j\varphi}e^{j\omega t}$$

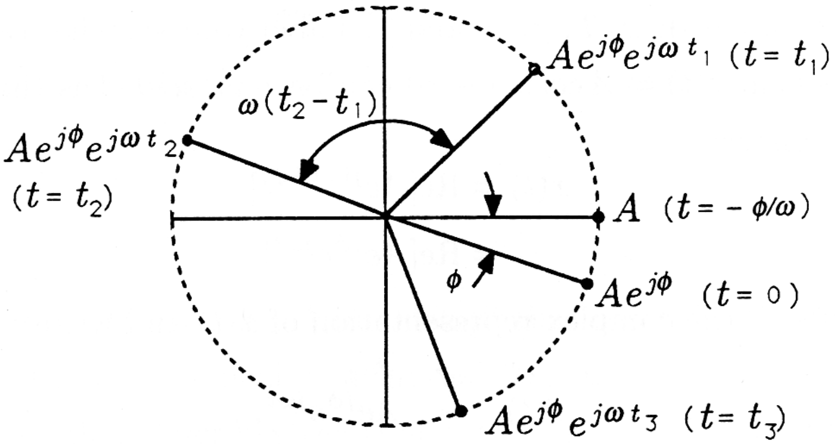
the complex representation of the real signal $A\cos\left(\omega t+\varphi\right)$. At $t=0$, the complex representation produces the *phasor*

Equation:

$$Ae^{j\varphi}.$$

This phasor is illustrated in [link]. In the figure, φ is approximately $\frac{-\pi}{10}$ If we let t increase to time t_1, then the complex representation produces the phasor

$$Ae^{j\phi}e^{j\omega t_1}.$$


$Ae^{j\phi}e^{j\omega t_1}$ $(t = t_1)$
$\omega(t_2 - t_1)$
$Ae^{j\phi}e^{j\omega t_2}$
$(t = t_2)$
A $(t = -\phi/\omega)$
ϕ
$Ae^{j\phi}$ $(t = 0)$
$Ae^{j\phi}e^{j\omega t_3}$ $(t = t_3)$

Rotating Phasor

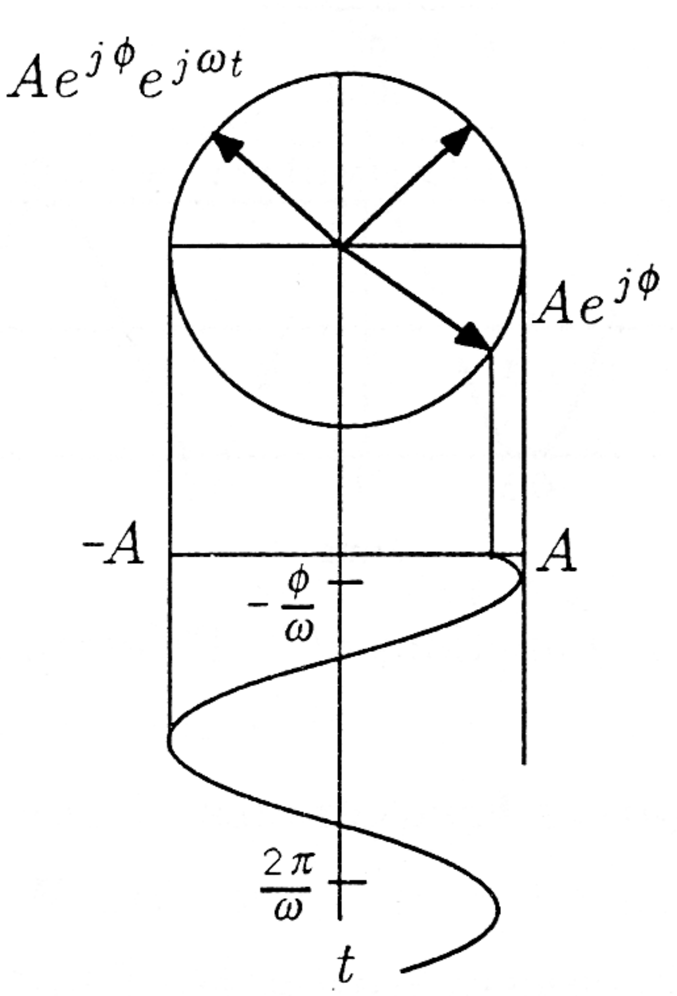
We know from our study of complex numbers that $e^{j\omega t_1}$ just rotates the phasor $Ae^{j\varphi}$ through an angle of ωt_1 ! See [link]. Therefore, as we run t from 0, indefinitely, we rotate the phasor $Ae^{j\varphi}$ indefinitely, turning out the circular trajectory of [link]. When $t = \frac{2\pi}{\omega}$ then $e^{j\omega t} = e^{j2\pi} = 1.$ Therefore, every $\left(\frac{2\pi}{\omega}\right)$ seconds, the phasor revisits any given position on the circle of radius A. We sometimes call $Ae^{j\varphi}e^{j\omega t}$ a *rotating phasor* whose rotation rate is the frequency ω:

Equation:

$$\frac{d}{dt}\omega t = \omega.$$

This rotation rate is also the frequency of the cosinusoidal signal $A\cos\left(\omega t + \varphi\right)$.

In summary, $Ae^{j\varphi}e^{j\omega t}$ is the complex, or rotating phasor, representation of the signal $A\cos\left(\omega t + \varphi\right)$. In this representation, $e^{j\omega t}$ rotates the phasor $Ae^{j\varphi}$ through angles ωt at the rate ω. The real part of the complex representation is the desired signal $A\text{cos}\left(\omega t + \varphi\right)$. This real part is read off the rotating phasor diagram as illustrated in [link]. In the figure, the angle φ is about $-\frac{2\pi}{10}$. As we become more facile with phasor representations, we will write $x\left(t\right) = \text{Re}\left[Xe^{j\omega t}\right]$ and call $Xe^{j\omega t}$ the complex representation and X the phasor representation. The phasor X is, of course, just the phasor $Ae^{j\varphi}$.



Reading a Real Signal from a Complex, Rotating Phasor

Exercise:

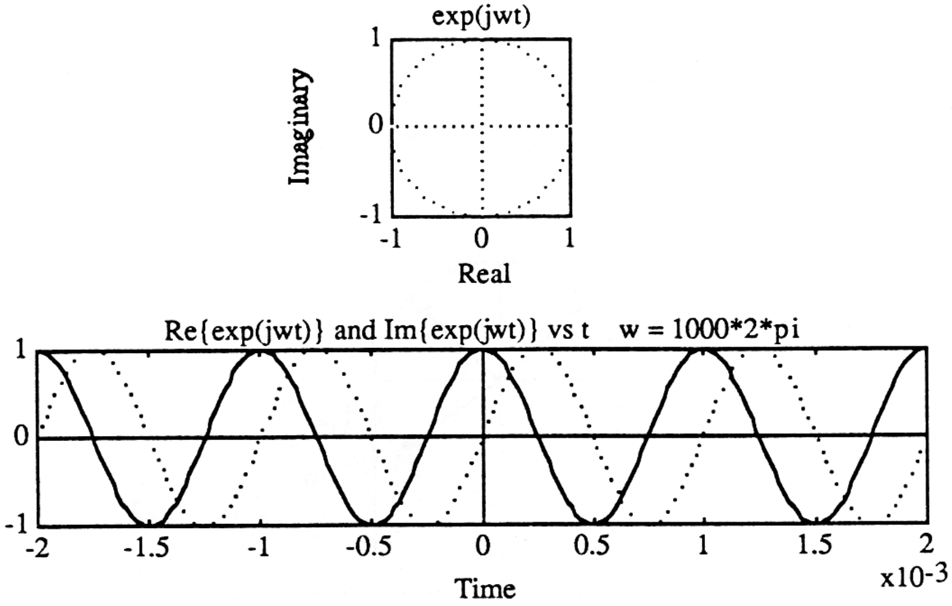
Problem:

Sketch the imaginary part of $Ae^{j\varphi}e^{j\omega t}$ to show that this is $A\ \sin\left(\omega t+\varphi\right)$. What do we mean when we say that the real and imaginary parts of $Ae^{j\varphi}e^{j\omega t}$ are "90° out of phase"?

Exercise:

Problem:

(MATLAB) Modify Demo 2.1 in "The Function e^x and $e^{j\theta}$" so that $\theta = \omega t$, with ω an input frequency variable and t a time variable that ranges from $-2\left(\frac{2\pi}{\omega}\right)$ to $+2\left(\frac{2\pi}{\omega}\right)$ in steps of 0.02 $\left(\frac{2\pi}{\omega}\right)$. In your modified program, compute and plot $e^{j\omega t}$, $\mathrm{Re}\left[e^{j\omega t}\right]$, and $\mathrm{Im}\left[e^{j\omega t}\right]$ for $-2\left(\frac{2\pi}{\omega}\right) \le t < 2\left(\frac{2\pi}{\omega}\right)$ in steps of 0.02 $\left(\frac{2\pi}{\omega}\right)$. Plot $e^{j\omega t}$ in a two-dimensional plot to get a picture like [link] and plot $\mathrm{Re}\left[e^{j\omega t}\right]$ and $\mathrm{Im}\left[e^{j\omega t}\right]$ versus t to get signals like those of [link]. You should observe something like [link] using the subplot features discussed in An Introduction to MATLAB. (In the figure, w represents Greek ω.)



The Functions $e^{j\omega t}$, $Re\left[e^{j\omega t}\right]$, and $Im\left[e^{j\omega t}\right]$

Positive and Negative Frequencies. There is an alternative phasor representation for the signal $x(t) = A\cos(\omega t + \varphi)$. We obtain it by using the Euler formula of "The Function e^x $s^{j\theta}$", namely, $\cos\theta = \frac{1}{2}\left(e^{j\theta} + e^{-j\theta}\right)$. When this formula is applied to $x(t)$, we obtain the result

Equation:

$$\begin{aligned} x\left(t\right) &= \tfrac{A}{2}\left[e^{j(\omega t+\varphi)} + e^{-j(\omega t+\varphi)}\right] \\ &= \tfrac{A}{2}e^{j\varphi}e^{j\omega t} + \tfrac{A}{2}e^{-j\varphi}e^{-j\omega t}. \end{aligned}$$

In this formula, the term $\frac{A}{2}e^{j\varphi}e^{j\omega t}$ is a rotating phasor that begins at the phasor value $Ae^{j\varphi}$ (for $t = 0$) and rotates counterclockwise with frequency ω. The term $\frac{A}{2}e^{-j\varphi}e^{-j\omega t}$ is a rotating phasor that begins at the (complex conjugate) phasor value $\frac{A}{2}e^{-j\varphi}$ (for $t = 0$) and rotates clockwise with (negative) ffequency ω. The physically meaningful frequency for a cosine is ω, a positive number like $2\pi(60)$ for 60 Hz power. There is no such thing as a negative frequency. The so-called negative frequency of the term $\frac{A}{2}e^{-j\varphi}e^{-j\omega t}$ just indicates that the direction of rotation for the rotating phasor is clock-wise and not counterclockwise. The notion of a negative frequency is just an artifact of the two-phasor representation of A $\cos\left(\omega t + \varphi\right)$. In the one-phasor representation, when we take the “real part,” the artifact does not arise. In your study of circuits, systems theory, electromagnetics, solid-state devices, signal processing, control, and communications, you will encounter both the one- and two-phasor representations. Become facile with them.

Exercise:

Problem:

Sketch the two-phasor representation of A $\cos\left(\omega t + \varphi\right)$. Show clearly how this representation works by discussing the counterclockwise rotation of the positive frequency part and the clockwise rotation of the negative frequency part.

Adding Phasors. The sum of two signals with common frequencies but different amplitudes and phases is

Equation:

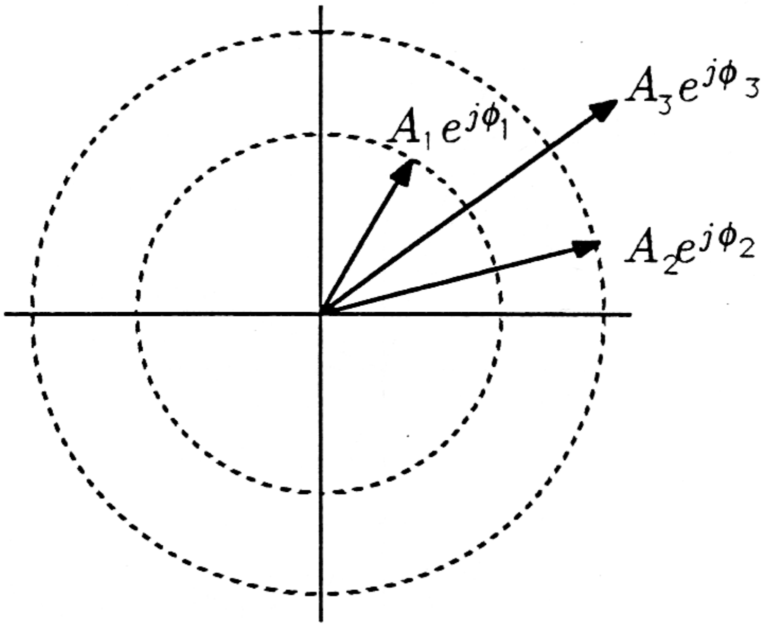
$$A_1 \cos\left(\omega t + \varphi_1\right) + A_2 \cos\left(\omega t + \varphi_2\right).$$

The rotating phasor representation for this sum is

Equation:

$$\left(A_1 e^{j\varphi_1} + A_2 e^{j\varphi_2}\right) e^{j\omega t}.$$

The new phasor is $A_1 e^{j\varphi_1} + A_2 e^{j\varphi_2}$, and the corresponding real signal is $x\left(t\right) = \mathrm{Re}\left[\left(A_1 e^{j\varphi_1} + A_2 e^{j\varphi_2}\right) e^{j\omega t}\right]$. The new phasor is illustrated in [link].



Adding Phasors

Exercise:

Problem:

Write the phasor $A_1 e^{j\varphi_1} + A_2 e^{j\varphi_2}$ as $A_3 e^{j\varphi_3}$; determine A_3 and φ_3 in terms of A_1, A_2, φ_1, and φ_2. What is the corresponding real signal?

Differentiating and Integrating Phasors. The derivative of the signal $A\ \cos\ (\omega t + \varphi)$ is the signal

Equation:

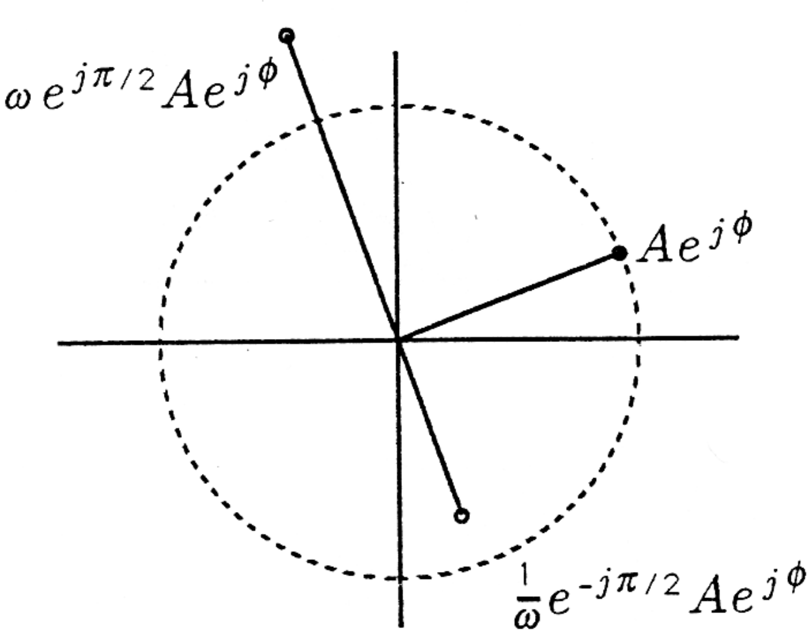
$$\begin{aligned} \frac{d}{dt}\ A\ \ \cos\ (\omega t + \varphi) &= -\omega A \sin\ (\omega t + \varphi) \\ &= -\operatorname{Im}\left[\omega A e^{j\varphi} e^{j\omega t}\right] \\ &= \operatorname{Re}\left[j\omega A e^{j\varphi} e^{j\omega t}\right] \\ &= \operatorname{Re}\left[\omega e^{j\pi/2} A e^{j\varphi} e^{j\omega t}\right]. \end{aligned}$$

This finding is very important. It says that the derivative of $A\ \cos\ (\omega t + \varphi)$ has the phasor representation

Equation:

$$\begin{aligned} \frac{d}{dt} A\ \cos\ \left(\omega t + \varphi\right) &\leftrightarrow j\omega A e^{j\varphi} \\ &\leftrightarrow \omega e^{j\pi/2} A e^{j\varphi}. \end{aligned}$$

These two phasor representations are entirely equivalent. The first says that the phasor $A e^{j\varphi}$ is complex scaled by $j\omega$ to produce the phasor for $\frac{d}{dt} A\ \cos\ (\omega t + \varphi)$, and the second says that it is scaled by ω and phased by $+\pi/2$. The phasor representations of $A\ \cos\ (\omega t + \varphi)$ and $\frac{d}{dt} A\ \cos\ (\omega t + \varphi)$ are illustrated in Figure 3.6. Note that the derivative “leads by $\pi/2$ radians (90)."



$\omega e^{j\pi/2} A e^{j\phi}$
$A e^{j\phi}$
$\frac{1}{\omega} e^{-j\pi/2} A e^{j\phi}$

Differentiating and Integrating Phasors

The integral of $A \cos(\omega t + \varphi)$ is
Equation:

$$\begin{aligned}\int A \cos(\omega t+\varphi)dt &= \frac{A}{\omega}\sin(\omega t+\varphi)\\ &= \mathrm{Im}\left[\frac{A}{\omega}e^{j\varphi}e^{j\omega t}\right]\\ &= \mathrm{Re}\left[-j\frac{A}{\omega}e^{j\varphi}e^{j\omega t}\right]\\ &= \mathrm{Re}\left[\frac{A}{j\omega}e^{j\varphi}e^{j\omega t}\right]\\ &= \mathrm{Re}\left[\frac{1}{\omega}e^{-j\pi/2}Ae^{j\varphi}e^{j\omega t}\right].\end{aligned}$$

This finding shows that the integral of $A \cos(\omega t + \varphi)$ has the phasor representation
Equation:

$$\begin{aligned}\int A\cos(\omega t+\varphi)dt &\leftrightarrow \frac{1}{j\omega}Ae^{j\varphi}\\ &\leftrightarrow \frac{1}{\omega}e^{-j\pi/2}Ae^{j\varphi}.\end{aligned}$$

The phasor $Ae^{j\varphi}$ is complex scaled by $\frac{1}{j\omega}$ or scaled by $\frac{1}{\omega}$ and phased by $e^{-j\pi/2}$ to produce the phasor for $\int A\cos\left(\omega t+\varphi\right)dt$. This is illustrated in [link]. Note that the integral “lags by $\pi/2$ radians (90). Keep these geometrical pictures of leading and lagging by $\pi/2$ in your mind at all times as you continue your more advanced study of engineering.

An Aside: The Harmonic Oscillator. The signal $A\cos\left(\omega t+\varphi\right)$ stands on its own as an interesting signal. But the fact that it reproduces itself (with scaling and phasing) under differentiation means that it obeys the second-order differential equation of the simple harmonic oscillator.[footnote] That is, the differential equation
This means, also, that we have an easy way to synthesize cosines with circuits that obey the equation of a simple harmonic oscillator!
Equation:

$$\frac{d^2x\left(t\right)}{dt^2}+\omega^2x\left(t\right)=0$$

has the solution
Equation:

$$x(t)=A\cos\left(\omega t+\varphi\right).$$

Try it:
Equation:

$$\frac{d^2}{dt^2}x\left(t\right)=\frac{d}{dt}[-A\omega\sin\left(\omega t+\varphi\right)]=-\omega^2A\cos\left(\omega t+\varphi\right).$$

The constants A and φ are determined from the initial conditions

Equation:

$$\begin{array}{ccc} x\left(0\right) = A\cos\varphi & & x^2\left(0\right) + x^2\left(\frac{\pi}{2\omega}\right) = A^2 \\ & \Leftrightarrow & \\ x\left(\frac{\pi}{2\omega}\right) = -A\sin\varphi & & -\frac{x(\pi/2\omega)}{x(0)} = \tan\varphi. \end{array}$$

Exercise:

Problem:

Show how to compute A and φ in the equation $x(t) = A\cos\left(\omega t + \varphi\right)$ from the initial conditions $x(\mathrm{O})$ and $\frac{d}{dt}x\left(t\right)|_{t=0}$.

Phasors: Introduction

Note: This module is part of the collection, *A First Course in Electrical and Computer Engineering*. The LaTeX source files for this collection were created using an optical character recognition technology, and because of this process there may be more errors than usual. Please contact us if you discover any errors.

Notes to Teachers and Students:

Phasors! For those who understand them, they are of incomparable value for the study of elementary and advanced topics. For those who misunderstand them, they are a constant source of confusion and are of no apparent use. So, let's understand them.

The conceptual leap from the complex number $e^{j\theta}$ to the phasor $e^{j(\omega t+\theta)}$ comes in "Phasor Representation of Signals". Take as long as necessary to understand every geometrical and algebraic nuance. Write the MATLAB program in "Exercise 6" to fix the key ideas once and for all. Then *use* phasors to study beating between tones, multiphase power, and Lissajous figures in "Beating between Tones" through "Lissajous Figures". We usually conduct a classroom demonstration of beating between tones using two phase-locked sources, an oscilloscope, and a speaker. We also demonstrate Lissajous figures with this hardware.

"Sinusoidal Steady State and the Series RLC Circuit" and Light Scattering by a Slit" on sinusoidal steady state and light scattering are too demanding for freshmen but are right on target for sophomores. These sections may be covered in a sophomore course (or a supplement to a sophomore course) or skipped in a freshman course without consequence.

In the numerical experiment in "Numerical Experiment (Interference Patterns)", students compute and plot interference patterns for two sinusoids that are out of phase.

Introduction

In engineering and applied science, three test signals form the basis for our study of electrical and mechanical systems. The *impulse* is an idealized signal that models very short excitations (like current pulses, hammer blows, pile drives, and light flashes). The *step* is an idealized signal that models excitations that are switched on and stay on (like current in a relay that closes or a transistor that switches). The *sinusoid* is an idealized signal that models excitations that oscillate with a regular frequency (like AC power, AM radio, pure musical tones, and harmonic vibrations). All three signals are used in the laboratory to design and analyze electrical and mechanical circuits, control systems, radio antennas, and the like. The sinusoidal signal is particularly important because it may be used to determine the frequency selectivity of a circuit (like a superheterodyne radio receiver) to excitations of different frequencies. For this reason, every manufacturer of electronics test equipment builds sinusoidal oscillators that may be swept through many octaves of Orequency. (Hewlett-Packard was started in 1940 with the famous HP audio oscillator.)

In this chapter we use what we have learned about complex numbers and the function $e^{j\theta}$ to develop a *phasor calculus* for representing and manipulating sinusoids. This calculus operates very much like the calculus we developed in "Complex Numbers" and "The Functions e^{x} and $e^{j\theta}$" for manipulating complex numbers. We ap- ply our calculus to the study of beating phenomena, multiphase power, series *RLC* circuits, and light scattering by a slit.

Phasors: Beating between Tones

Note: This module is part of the collection, *A First Course in Electrical and Computer Engineering*. The LaTeX source files for this collection were created using an optical character recognition technology, and because of this process there may be more errors than usual. Please contact us if you discover any errors.

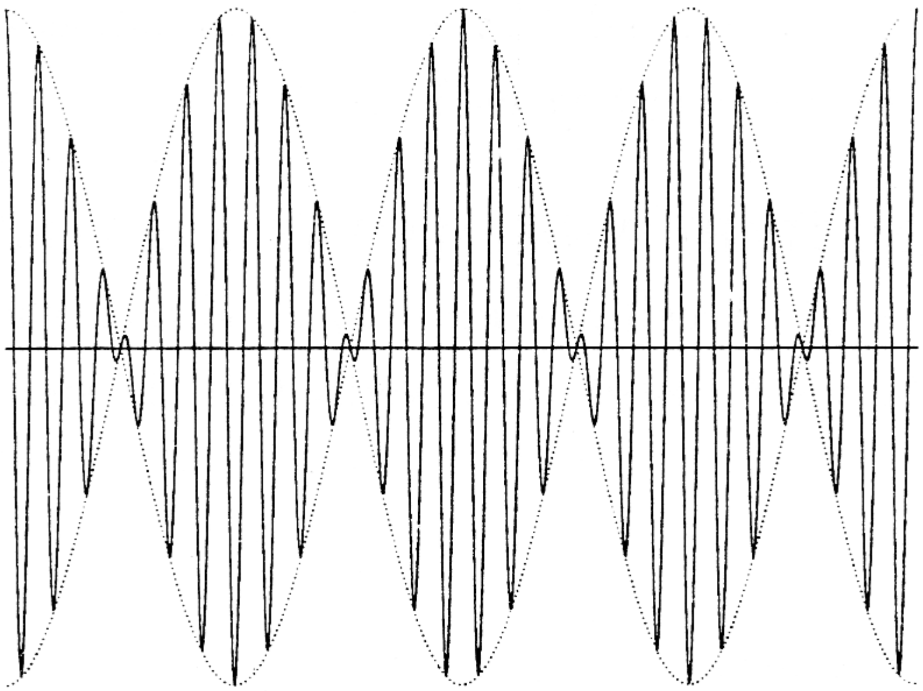
Perhaps you have heard two slightly mistuned musical instruments play pure tones whose frequencies are close but not equal. If so, you have sensed a beating phenomenon wherein a pure tone seems to wax and wane. This waxing and waning tone is, in fact, a tone whose frequency is the average of the two mismatched frequencies, amplitude modulated by a tone whose “beat” frequency is half the difference between the two mismatched frequencies. The effect is illustrated in [link]. Let's see if we can derive a mathematical model for the beating of tones.

We begin with two pure tones whose frequencies are $\omega_0 + \nu$ and $\omega_0 - \nu$ (for example, $\omega_0 = 2\pi \times 10^3 \mathrm{rad}/\ \mathrm{sec}$ and $\nu = 2\pi\ \mathrm{rad}\ /\ \mathrm{sec}$). The average frequency is ω_0, and the difference frequency is 2ν. What you hear is the sum of the two tones:
Equation:

$$x\left(t\right) = A_1 \cos\left[(\omega_0 + \nu)t + \varphi_1\right] + A_2 \cos\left[(\omega_0 - \nu)t + \varphi_2\right].$$

The first tone has amplitude A_1 and phase φ_1 ; the second has amplitude A_2 and phase φ_2. We will assume that the two amplitudes are equal to A. Furthermore, whatever the phases, we may write them as
Equation:

$$\begin{array}{ccc} \varphi_1 = \varphi + \psi & \text{and} & \varphi_2 = \varphi - \psi \\ \varphi = \frac{1}{2}(\varphi_1 + \varphi_2) & \text{and} & \psi = \frac{1}{2}(\varphi_1 - \varphi_2). \end{array}$$



Beating between Tones

Recall our trick for representing $x(t)$ as a complex phasor:
Equation:

$$\begin{aligned} x\,(t) &= A\ \mathrm{Re}\Big\{e^{j[(\omega_0+\nu)t+\varphi+\psi]}, +, e^{j[(\omega_0-\nu)t+\varphi-\psi]}\Big\} \\ &= A\ \mathrm{Re}\Big\{e^{j(\omega_0 t+\varphi)}, \Big[e^{j(\nu t+\psi)} + e^{-j(\nu t+\psi)}\Big]\Big\} \\ &= 2A\ \mathrm{Re}\Big\{e^{j(\omega_0 t+\varphi)}, \cos, (\nu t+\psi)\Big\} \\ &= 2A\ \cos\,(\omega_0 t+\varphi)\,\cos\,(\nu t+\psi). \end{aligned}$$

This is an amplitude modulated wave, wherein a low frequency signal with beat frequency ν rad/sec modulates a high frequency signal with carrier frequency ω_0 rad/sec. Over short periods of time, the modulating term $\cos\,(\nu t{+}\psi)$ remains essentially constant while the carrier term $\cos\,(\omega_0 t+\varphi)$ turns out many cycles of its tone. For example, if t runs from 0 to $\frac{2\pi}{10\nu}$ (about 0.1 seconds in our example), then the modulating wave turns out just 1/10 cycle while the carrier turns out $10\nu\omega_{\Delta}$ cycles (about 100 in our example). Every time νt changes by 2π radians, then the modulating term goes from a maximum (a wax) through a minimum (a wane) and back to a maximum. This cycle takes

Equation:

$$\nu t = 2\pi \Leftrightarrow t = \frac{2\pi}{\nu}\text{seconds},$$

which is 1 second in our example. In this 1 second the carrier turns out 1000 cycles.

Exercise:

Problem:

Find out the frequency of A above middle C on a piano. Assume two pianos are mistuned by $\pm 1\text{Hz}(\pm 2\pi\text{rad}/\text{sec})$. Find their beat frequency ν and their carrier frequency ω_0.

Exercise:

Problem: (MATLAB) Write a MATLAB program to compute and plot

A cos $[(\omega_0 + \nu)t + \varphi_1]$, A cos $[(\omega_0 - \nu)t + \varphi_2]$, and their sum. Then compute and plot $2A$ cos $(\omega_0 t + \varphi)$ cos $(\nu t + \psi)$.

Verify that the sum equals this latter signal.

Phasors: Multiphase Power

Note: This module is part of the collection, *A First Course in Electrical and Computer Engineering*. The LaTeX source files for this collection were created using an optical character recognition technology, and because of this process there may be more errors than usual. Please contact us if you discover any errors.

The electrical service to your home is a two-phase service.[footnote] This means that two 110 volt, 60 Hz lines, plus neutral, terminate in the panel. The lines are π radians (180) out of phase, so we can write them as
It really is, although it is said to be "single phase" because of the way it is picked off a single phase of a primary source. You will hear more about this in circuits and power courses.
Equation:

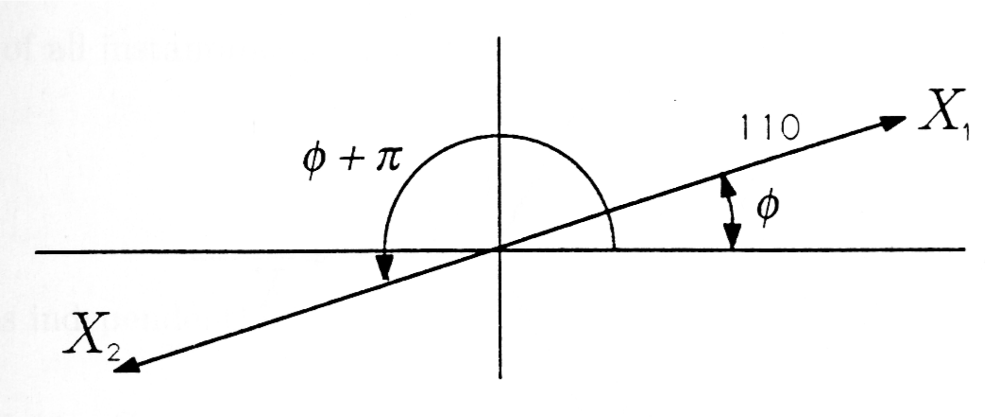
$$\begin{aligned} x_1(t) &= 110\ cos\ [2\pi(60)t+\varphi] = \mathrm{Re}\left\{110e^{j[2\pi(60)t+\varphi]}\right\} \\ &= \mathrm{Re}\left\{X_1 e^{j2\pi(60)t}\right\} \end{aligned}$$

$$X_1 = 110e^{j\varphi}$$

$$\begin{aligned} x_2(t) &= 110\ cos\ [2\pi(60)t+\varphi+\pi] = \mathrm{Re}\left\{110e^{j[2\pi(60)t+\varphi+\pi]}\right\} \\ &= \mathrm{Re}\left\{X_2 e^{j2\pi(60)t}\right\} \end{aligned}$$

$$X_2 = 110e^{j(\varphi+\pi)}.$$

These two voltages are illustrated as the phasors X_1 and X_2 in [link].



Phasors in Two-Phase Power

You may use $x_1\,(t)$ to drive your clock radio or your toaster and the difference between $x_1\,(t)$ and $x_2\,(t)$ to drive your range or dryer:

Equation:

$$x_1\,(t) - x_2\,(t) = 220\ \cos\,[2\pi(60)t + \varphi].$$

The phasor representation of this difference is

Equation:

$$X_1 - X_2 = 220e^{j\varphi}.$$

The breakers in a breaker box span the x_1-to-neutral bus for 110 volts and the x_1-to-x_2 buses for 220 volts.

Exercise:

Problem: Sketch the phasor $X_1 - X_2$ on [link].

Most industrial installations use a three-phase service consisting of the signals $x_1\,(t)$, $x_2\,(t)$, and $x_3\,(t)$:

Equation:

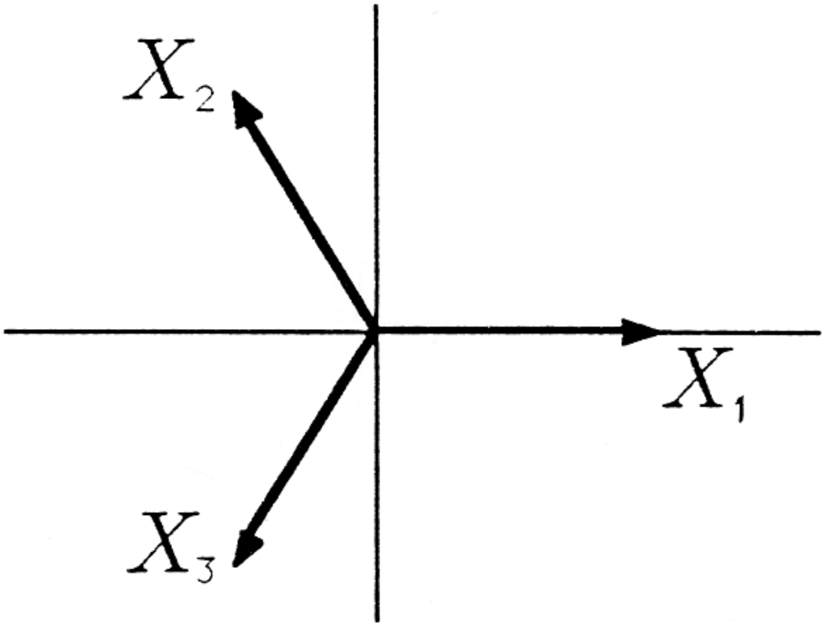
$$x_n(t) = 110\,\mathrm{Re}\left\{e^{j[\omega_0 t + n(2\pi/3)]}\right\} \leftrightarrow X_n = 110 e^{jn(2\pi/3)}, \;\; n = 1, 2, 3.$$

The phasors for three-phase power are illustrated in [link].

Exercise:

Problem:

Sketch the phasor $X_2 - X_1$ corresponding to $x_2(t) - x_1(t)$ on [link]. Compute the voltage you can get with $x_2(t) - x_1(t)$. This answer explains why you do not get 220 volts in three-phase circuits. What do you get?



Three-Phase Power

Constant Power. Two- and three-phase power generalizes in an obvious way to N-phase power. In such a scheme, the N signals $x_n(n = 0, 1, \ldots, N - 1)$ are

Equation:

$$\begin{aligned} x_n(t) &= A\cos\left(\omega t + \frac{2\pi}{N}n\right) \\ &= \mathrm{Re}\left[Ae^{j2\pi n/N}e^{j\omega t}\right] \leftrightarrow X_n = Ae^{j2\pi n/N}. \end{aligned}$$

The phasors X_n are $Ae^{j2\pi(n/N)}$. The sum of all N signals is zero:

Equation:

$$\begin{aligned} \textstyle\sum_{n=0}^{N-1} x_n(t) &= \mathrm{Re}\left\{A\textstyle\sum_{n=0}^{N-1} e^{j2\pi n/N}e^{j\omega t}\right\} \\ &= \mathrm{Re}\left\{A\frac{1-e^{j2\pi}}{1-e^{j2\pi/N}}e^{j\omega t}\right\} \\ &= 0. \end{aligned}$$

But what about the sum of the instantaneous powers? Define the instantaneous power of the n^{th} signal to be

Equation:

$$\begin{aligned} p_n(t) = x_n^2(t) &= A^2\cos^2\left(\omega t + \tfrac{2\pi}{N}n\right) \\ &= \tfrac{A^2}{2} + \tfrac{A^2}{2}\cos\left(2\omega t + 2\tfrac{2\pi}{N}n\right) \\ &= \tfrac{A^2}{2} + \mathrm{Re}\left\{\tfrac{A^2}{2}e^{j(2\pi/N)2n}e^{j2\omega t}\right\}. \end{aligned}$$

The sum of all instantaneous powers is (see [link])

Equation:

$$P = \sum_{n=0}^{N-1} p_n(t) = N\frac{A^2}{2},$$

and this is independent of time!

Exercise:

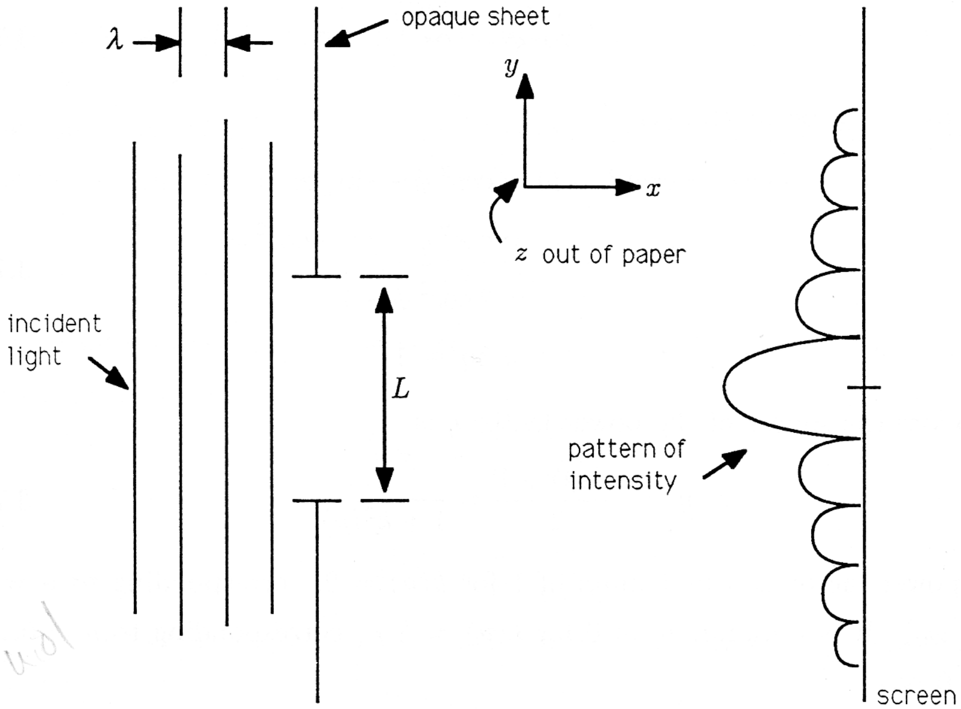
Problem:

Carry out the computations of Equation 3.33 to prove that instantaneous power P is constant in the N-phase power scheme.

Phasors: Light Scattering by a Slit

Note: This module is part of the collection, *A First Course in Electrical and Computer Engineering*. The LaTeX source files for this collection were created using an optical character recognition technology, and because of this process there may be more errors than usual. Please contact us if you discover any errors.

One of the most spectacular successes for phasor analysis arises in the study of light diffraction by a narrow slit. The experiment is to shine laser light through a slit in an otherwise opaque sheet and observe the pattern of light that falls on a distant screen. When the slit is very narrow compared with the wavelength of the light, then the light that falls on the screen is nearly uniform in intensity. However, when the width of the slit is comparable to the wavelength of the light, the pattern of light that falls on the screen is scalloped in intensity, showing alternating light and dark bands. The experiment, and the observed results, are illustrated in [link].



Light Diffraction by a Slit

Why should this experiment produce this result? Phasor analysis illuminates the question and produces an elegant mathematical description of a profoundly important optical experiment.

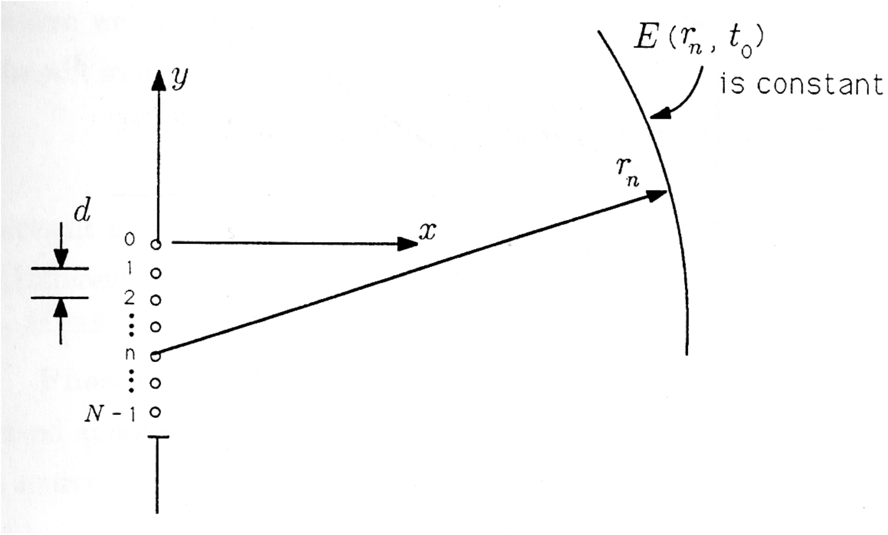
Huygens's Principle. We will assume, as Christiaan Huygens did, that the light incident on the slit sets up a light field in the slit that may be modeled by N discrete sources, each of which radiates a “spherical wave of light.” This model is illustrated in [link]. The distance between sources is d, and $Nd = L$ is the width of the slit. Each source is indexed by n, and n runs from 0 to $N - 1$. The 0^{th} source is located at the origin of our coordinate system.

The spherical wave radiated by the n^{th} source is described by the equation

Equation:

$$E\left(r_n, t\right) = \text{Re}\left\{\frac{A}{N} e^{j[\omega t - (2\pi/\lambda) r_n]}\right\}.$$

The function $E(r_n, t)$ describes the “electric field” at time t and distance r_n from the n^{th} source. The field is constant as long as the variable $\omega t - (2\pi/\lambda) r_n$ is constant. Therefore, if we freeze time at $t = t_0$, the field will be constant on a sphere of radius r_n. This is illustrated in [link].



Huygens's Model for Light Diffraction

Exercise:

Problem:

Fix r_n in $E(r_n, t)$ and show that $E(r_n, t)$ varies cosinusoidally with time t. Sketch the function and interpret it. What is its period?

Exercise:

Problem:

Fix t in $E(r_n, t)$ and show that $E(r_n, t)$ varies cosinusoidally with radius r_n. Sketch the function and interpret it. Call the “period in r_n" the “wavelength.” Show that the wavelength is λ.

Exercise:

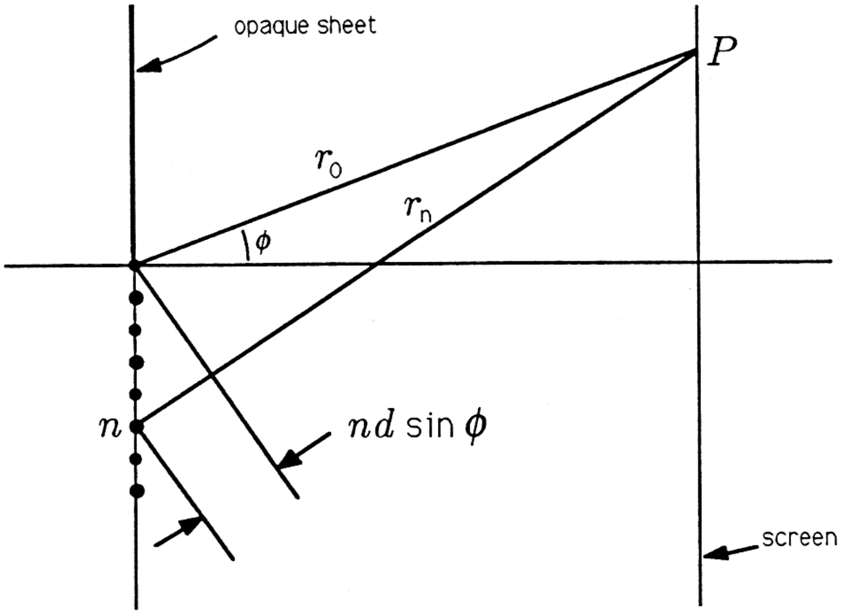
Problem:

The “crest of the wave $E(r_n, t)$" occurs when $\omega t - (2\pi/\lambda)r_n = 0$. Show that the crest moves through space at velocity $v = \omega\lambda/2\pi$.

Geometry. If we now pick a point P on a distant screen, that point will be at distance r_0 from source 0, ..., r_n from source n, ..., and so on. If we isolate the sources 0 and n, then we have the geometric picture of [link]. The angle φ is the angle that point P makes with the horizontal axis. The Pythagorean theorem says that the connection between distances r_0 and r_n is

Equation:

$$(r_n - nd \sin \varphi)^2 + (nd \cos \varphi)^2 = r_0^2.$$



Geometry of the Experiment

Let's try the solution

Equation:

$$r_n = r_0 + nd \sin \varphi.$$

This solution produces the approximate identity

Equation:

$$r_0^2 + (nd \cos \varphi)^2 \cong r_0^2$$

Equation:

$$r_0^2 + (nd\cos\varphi)^2 \cong r_0^2$$
$$1 + \left(\frac{nd}{r_0}\ cos\ \varphi\right)^2 \cong 1.$$

This will be close for $\frac{nd}{r_0} << 1$. We will assume that the slit width L is small compared to the distance to any point on the screen. Then $\frac{Nd}{r_0} = \frac{L}{r_0} << 1$, in which case the approximate solution for r_n is valid for all n. This means that, for any point P on the distant screen, the light contributed by the n^{th} source is approximately

Equation:

$$\begin{aligned} E_n(\varphi, t) &= \mathrm{Re}\left\{\frac{A}{N} e^{j[\omega t - (2\pi/\lambda)(r_0 + nd\sin\varphi)]}\right\} \\ &= \mathrm{Re}\left\{\frac{A}{N} e^{-j(2\pi/\lambda)r_0} e^{-j(2\pi/\lambda)nd\sin\varphi} e^{j\omega t}\right\}. \end{aligned}$$

The phasor representation for this function is just

Equation:

$$E_n(\varphi) = \frac{A}{N} e^{-j(2\pi/\lambda)r_0} e^{-j(2\pi/\lambda)nd\sin\varphi}.$$

Note that $E_0(\varphi)$, the phasor associated with the 0^{th} source, is $\frac{A}{N} e^{-j(2\pi/\lambda)r_0}$. Therefore we may write the phasor representation for the light contributed by the n^{th} source to be

Equation:

$$E_n(\varphi) = E_0(\varphi) e^{-j(2\pi/\lambda)nd\sin\varphi}.$$

This result is very important because it shows the light arriving at point P from different sources to be “out of phase” by an amount that depends on the ratio $\frac{nd\,\sin\varphi}{\lambda}$

Phasors and Interference. The phasor representation for the field observed at point P on the screen is the sum of the phasors contributed by each source:

Equation:

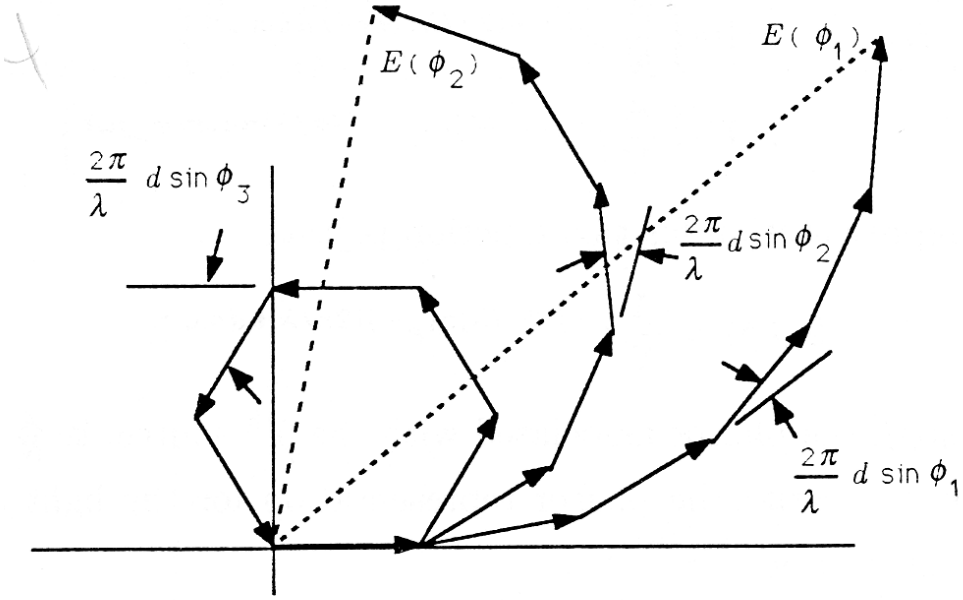
$$E\left(\varphi\right)=\sum_{n=0}^{N-1}E_n\left(\varphi\right)=E_0\left(\varphi\right)\sum_{n=0}^{N-1}e^{-j(2\pi/\lambda)nd\sin\varphi}.$$

This is a sum of the form

Equation:

$$E\left(\varphi\right)=E_0\left(\varphi\right)\sum_{n=0}^{N-1}e^{jn\theta}$$

where the angle θ is $(2\pi/\lambda)d\,\sin\,\varphi$. This sum is illustrated in [link] for several representative values of θ. Note that for small θ, meaning small φ, the sum has large magnitude, whereas for θ on the order of $2\pi/N$, the sum is small. This simple geometric interpretation shows that for some values of φ, corresponding to some points P on the screen, there will be constructive interference between the phasors, while for other values of φ there will be destructive interference. Constructive interference produces bright light, and destructive interference produces darkness.


$E(\phi_2)$
$E(\phi_1)$
$\frac{2\pi}{\lambda}\, d \sin \phi_3$
$\frac{2\pi}{\lambda}\, d \sin \phi_2$
$\frac{2\pi}{\lambda}\, d \sin \phi_1$

Phasor Sums for Diffraction

The geometry of [link] is illuminating. However, we already know from our study of complex numbers and geometric sums that the phasor sum of Equation 3.85 may be written as

Equation:

$$E(\varphi) = E_0(\varphi)\frac{1 - e^{-j(2\pi/\lambda)N d \sin\varphi}}{1 - e^{-j(2\pi/\lambda) d \sin\varphi}}.$$

This result may be manipulated to produce the form

Equation:

$$E(\varphi) = \frac{A}{N} e^{-j(2\pi/\lambda)r_0} e^{-j(\pi/\lambda)(N-1) d \sin\varphi} \frac{\sin\left(N\pi\frac{d}{\lambda}\sin\varphi\right)}{\sin\left(\pi\frac{d}{\lambda}\sin\varphi\right)}$$

The magnitude is the intensity of the light at angle φ from horizontal:

Equation:

$$|E(\varphi)| = \left|\frac{A}{N}\frac{\sin\left(N\pi\frac{d}{\lambda}\sin\varphi\right)}{\sin\left(\pi\frac{d}{\lambda}\sin\varphi\right)}\right|.$$

Exercise:

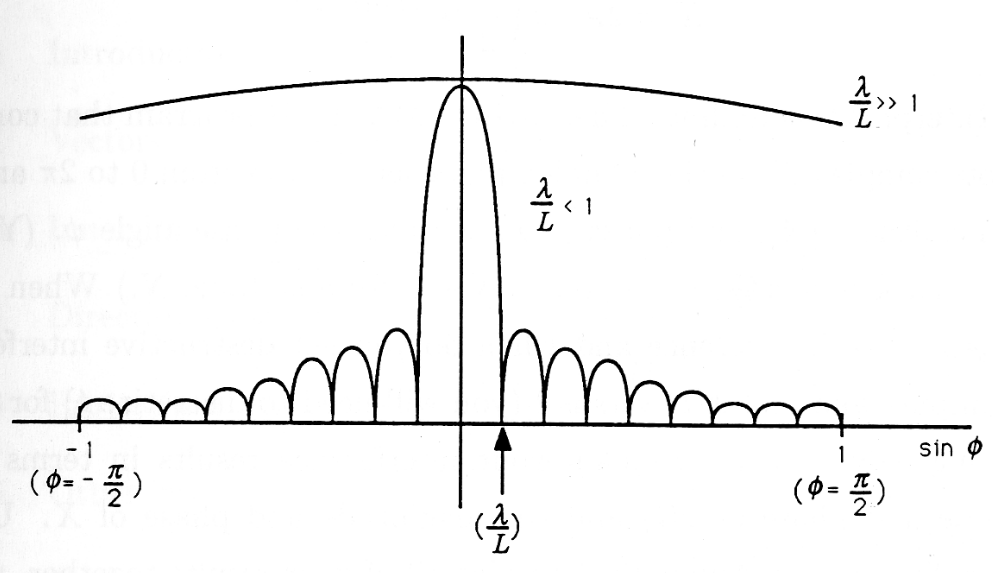
Problem: Derive [link] from [link].

Limiting Form. Huygens's model is exact when d shrinks to 0 and N increases to infinity in such a way that $Nd \to L$, the slit width. Then

Equation:

$$|E(\varphi)| \to \left|\frac{A\sin\left(\frac{\pi L}{\lambda}\sin\varphi\right)}{\frac{\pi L}{\lambda}\sin\varphi}\right|.$$

This function is plotted in [link] for two values of $\frac{L}{\lambda}$ the width of the slit measured in wavelengths. When $\frac{L}{\lambda} << 1$ (i.e., $\frac{\lambda}{L} >> 1$), then the light is uniformly distributed on the screen. However, when $\frac{L}{\lambda} > 1 \left(\frac{\lambda}{L} < 1\right)$, then the function has many zeros for $|\sin\varphi| < 1$, as illustrated in the figure. These zeros correspond to dark spots on the screen where the fields radiated from the infinity of points within the slit interfere destructively.



Interference Pattern for a Slit

Exercise:

Problem: Derive [link] from [link].

Exercise:

Problem: (MATLAB) Plot the discrete approximation
Equation:

$$\left|\frac{\frac{A}{N}\sin\left(N\pi\frac{d}{\lambda}\sin\varphi\right)}{\sin\left(\pi\frac{d}{\lambda}\sin\varphi\right)}\right|$$

versus $\sin\varphi$ for $\frac{L}{\lambda}=\frac{Nd}{\lambda}=10$ and $N=2,4,8,16,32$. Compare with the continuous, limiting form
Equation:

$$\left| \frac{A \sin\left(\pi \frac{L}{\lambda} \sin \varphi\right)}{\pi \frac{L}{\lambda} \sin \varphi} \right|$$

Fundamental Constants

Quantity	**Symbol**	**Value**
Atomic mass unit	u	$1.660\ 538\ 782\ \ (83)\ \times\ 10^{-27}\,\text{kg}$ $931.494\ 028\ \ (23)\,\text{MeV}/c^2$
Avogadro's number	N_A	$6.02214076\ \times\ 10^{23}\,\text{reciprocal mole}\left(\text{mol}^{-1}\right)$
Bohr magneton	$\mu_\text{B} = \frac{e\hbar}{2m_e}$	$9.274\ 009\ 15\ \ (23)\ \times\ 10^{-24}\,\text{J}/\text{T}$
Bohr radius	$a_0 = \frac{\hbar^2}{m_e e^2 k_e}$	$5.291\ 772\ 085\ 9\ \ (36)\ \times\ 10^{-11}\,\text{m}$
Boltzmann's constant	$k_\text{B} = \frac{R}{N_\text{A}}$	$1.380649\ \times\ 10^{-23}\,\text{joule per kelvin}\left(\text{J}\cdot\text{K}^{-1}\right)$
Compton wavelength	$\lambda_\text{C} = \frac{h}{m_e c}$	$2.426\ 310\ 217\ 5\ \ (33)\ \times\ 10^{-12}\,\text{m}$
Coulomb constant	$k_e = \frac{1}{4\pi\varepsilon_0}$	$8.987\ 551\ 788...\ \times\ 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2\ (\text{exact})$
Deuteron mass	m_d	$3.343\ 583\ 20\ \ (17)\ \times\ 10^{-27}\,\text{kg}$ $2.013\ 553\ 212\ 724\,(78)\,\text{u}$ $1875.612\ 859\,\text{MeV}/c^2$
Electron mass	m_e	$9.109\ 382\ 15\ \ (45)\ \times\ 10^{-31}\,\text{kg}$ $5.485\ 799\ 094\ 3\,(23)\ \times\ 10^{-4}\,\text{u}$ $0.510\ 998\ 910\ \ (13)\,\text{MeV}/c^2$
Electron volt	eV	$1.602\ 176\ 487\ \ (40)\ \times\ 10^{-19}\,\text{J}$

Quantity	**Symbol**	**Value**
Elementary charge	e	$1.602176634\ \times\ 10^{-19}\,\mathrm{C}$
Gas constant	R	$8.314\,472\ (15)\,\mathrm{J/mol\cdot K}$
Gravitational constant	G	$6.674\,28\ (67)\ \times\ 10^{-11}\,\mathrm{N\cdot m^2/kg^2}$
Neutron mass	m_n	$1.674\,927\,211\ (84)\ \times\ 10^{-27}\,\mathrm{kg}$ $1.008\,664\,915\,97\ (43)\,\mathrm{u}$ $939.565\,346\ (23)\,\mathrm{MeV}/c^2$
Nuclear magneton	$\mu_n=\frac{e\hbar}{2m_p}$	$5.050\,783\,24\ (13)\ \times\ 10^{-27}\,\mathrm{J/T}$
Permeability of free space	μ_0	$4\pi\ \times\ 10^{-7}\,\mathrm{T\cdot m/A}$ (exact)
Permittivity of free space	$\varepsilon_0=\frac{1}{\mu_0 c^2}$	$8.854\,187\,817...\ \times\ 10^{-12}\,\mathrm{C^2/N\cdot m^2}$ (exact)
Planck's constant	h $\hbar=\frac{h}{2\pi}$	$6.62607015\ \times\ 10^{-34}\,\mathrm{kg\cdot m^2\cdot s^{-1}}$ $1.05457182\ \times\ 10^{-34}\,\mathrm{kg\cdot m^2\cdot s^{-1}}$
Proton mass	m_p	$1.672\,621\,637\ (83)\ \times\ 10^{-27}\,\mathrm{kg}$ $1.007\,276\,466\,77\ (10)\,\mathrm{u}$ $938.272\,013\ (23)\,\mathrm{MeV}/c^2$
Rydberg constant	R_{H}	$1.097\,373\,156\,852\,7\ (73)\ \times\ 10^{7}\,\mathrm{m^{-1}}$
Speed of light in vacuum	c	$2.997\,924\,58\ \times\ 10^{8}\,\mathrm{m/s}$ (exact)

Fundamental Constants*Note:* These constants are the values recommended in 2006 by CODATA, based on a least-squares adjustment of data from different measurements. The numbers in parentheses for the values represent the uncertainties of the last two digits.

Useful combinations of constants for calculations:

$hc = 12{,}400\,\mathrm{eV}\cdot\text{Å} = 1240\,\mathrm{eV}\cdot\mathrm{nm} = 1240\,\mathrm{MeV}\cdot\mathrm{fm}$

$\hbar c = 1973\,\mathrm{eV}\cdot\text{Å} = 197.3\,\mathrm{eV}\cdot\mathrm{nm} = 197.3\,\mathrm{MeV}\cdot\mathrm{fm}$

$k_e e^2 = 14.40\,\mathrm{eV}\cdot\text{Å} = 1.440\,\mathrm{eV}\cdot\mathrm{nm} = 1.440\,\mathrm{MeV}\cdot\mathrm{fm}$

$k_\mathrm{B}T = 0.02585\,\mathrm{eV}\ \text{at}\ T = 300\,\mathrm{K}$

Mathematical Formulas

Quadratic formula

If $ax^2 + bx + c = 0$, then $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Triangle of base b and height h	**Area** $= \frac{1}{2}bh$	
Circle of radius r	Circumference $= 2\pi r$	Area $= \pi r^2$
Sphere of radius r	Surface area $= 4\pi r^2$	Volume $= \frac{4}{3}\pi r^3$
Cylinder of radius r and height h	Area of curved surface $= 2\pi rh$	Volume $= \pi r^2 h$

Geometry

Trigonometry

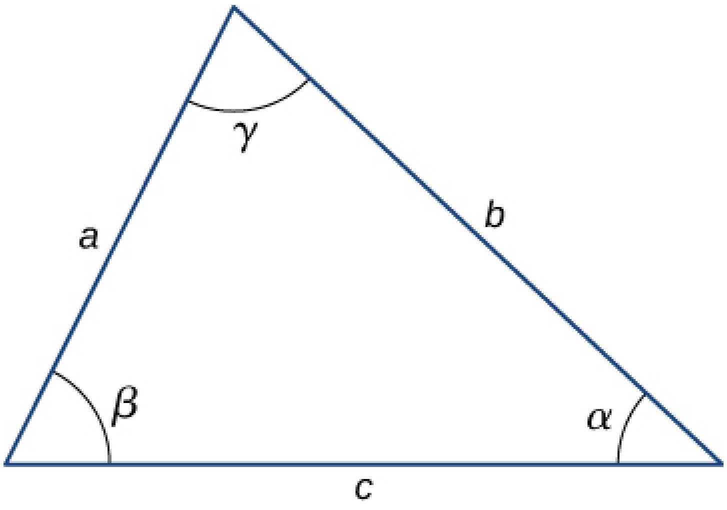
Trigonometric Identities

1. $\sin\theta = 1/\csc\theta$
2. $\cos\theta = 1/\sec\theta$
3. $\tan\theta = 1/\cot\theta$
4. $\sin\left(90^0 - \theta\right) = \cos\theta$
5. $\cos\left(90^0 - \theta\right) = \sin\theta$
6. $\tan\left(90^0 - \theta\right) = \cot\theta$
7. $\sin^2\theta + \cos^2\theta = 1$
8. $\sec^2\theta - \tan^2\theta = 1$

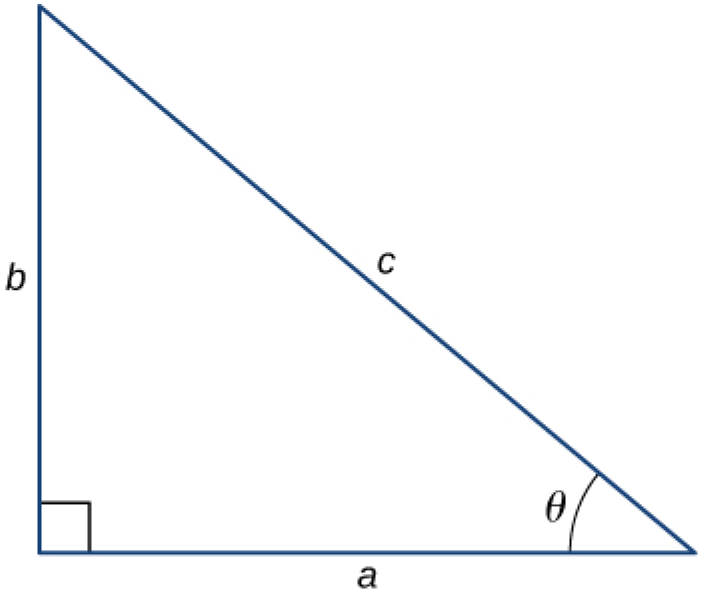
9. $\tan\theta = \sin\theta / \cos\theta$
10. $\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$
11. $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$
12. $\tan(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$
13. $\sin 2\theta = 2\sin\theta\cos\theta$
14. $\cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$
15. $\sin\alpha + \sin\beta = 2\sin\frac{1}{2}(\alpha + \beta)\cos\frac{1}{2}(\alpha - \beta)$
16. $\cos\alpha + \cos\beta = 2\cos\frac{1}{2}(\alpha + \beta)\cos\frac{1}{2}(\alpha - \beta)$

Triangles

1. Law of sines: $\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma}$
2. Law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\gamma$



3. Pythagorean theorem: $a^2 + b^2 = c^2$



Series expansions

1. Binomial theorem:
   $(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)a^{n-2}b^2}{2!} + \frac{n(n-1)(n-2)a^{n-3}b^3}{3!} + \cdots$
2. $(1 \pm x)^n = 1 \pm \frac{nx}{1!} + \frac{n(n-1)x^2}{2!} \pm \cdots \left(x^2 < 1\right)$
3. $(1 \pm x)^{-n} = 1 \mp \frac{nx}{1!} + \frac{n(n+1)x^2}{2!} \mp \cdots \left(x^2 < 1\right)$
4. $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$
5. $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$
6. $\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \cdots$
7. $e^x = 1 + x + \frac{x^2}{2!} + \cdots$
8. $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots (|x| < 1)$

Derivatives

1. $\frac{d}{dx}\left[af(x)\right] = a\frac{d}{dx}f(x)$
2. $\frac{d}{dx}\left[f(x) + g(x)\right] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$
3. $\frac{d}{dx}\left[f(x)g(x)\right] = f(x)\frac{d}{dx}g(x) + g(x)\frac{d}{dx}f(x)$
4. $\frac{d}{dx}f(u) = \left[\frac{d}{du}f(u)\right]\frac{du}{dx}$
5. $\frac{d}{dx}x^m = mx^{m-1}$
6. $\frac{d}{dx}\sin x = \cos x$

7. $\frac{d}{dx} \cos x = -\sin x$
8. $\frac{d}{dx} \tan x = \sec^2 x$
9. $\frac{d}{dx} \cot x = -\csc^2 x$
10. $\frac{d}{dx} \sec x = \tan x \sec x$
11. $\frac{d}{dx} \csc x = -\cot x \csc x$
12. $\frac{d}{dx} e^x = e^x$
13. $\frac{d}{dx} \ln x = \frac{1}{x}$
14. $\frac{d}{dx} \sin^{-1} x = \frac{1}{\sqrt{1-x^2}}$
15. $\frac{d}{dx} \cos^{-1} x = -\frac{1}{\sqrt{1-x^2}}$
16. $\frac{d}{dx} \tan^{-1} x = \frac{1}{1+x^2}$

Integrals

1. $\displaystyle\int af(x)dx = a\int f(x)dx$
2. $\displaystyle\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$
3. $\displaystyle\begin{aligned}\int x^m dx &= \frac{x^{m+1}}{m+1}\ (m \neq -1)\\ &= \ln x (m = -1)\end{aligned}$
4. $\displaystyle\int \sin x\, dx = -\cos x$
5. $\displaystyle\int \cos x\, dx = \sin x$
6. $\displaystyle\int \tan x\, dx = \ln|\sec x|$
7. $\displaystyle\int \sin^2 ax\, dx = \frac{x}{2} - \frac{\sin 2ax}{4a}$
8. $\displaystyle\int \cos^2 ax\, dx = \frac{x}{2} + \frac{\sin 2ax}{4a}$
9. $\displaystyle\int \sin ax \cos ax\, dx = -\frac{\cos 2ax}{4a}$

10. $\int e^{ax}\,dx = \frac{1}{a}e^{ax}$

11. $\int xe^{ax}dx = \frac{e^{ax}}{a^2}(ax-1)$

12. $\int \ln ax\,dx = x\ln ax - x$

13. $\int \dfrac{dx}{a^2+x^2} = \frac{1}{a}\tan^{-1}\frac{x}{a}$

14. $\int \dfrac{dx}{a^2-x^2} = \frac{1}{2a}\ln\left|\frac{x+a}{x-a}\right|$

15. $\int \dfrac{dx}{\sqrt{a^2+x^2}} = \sinh^{-1}\frac{x}{a}$

16. $\int \dfrac{dx}{\sqrt{a^2-x^2}} = \sin^{-1}\frac{x}{a}$

17. $\int \sqrt{a^2+x^2}\,dx = \frac{x}{2}\sqrt{a^2+x^2} + \frac{a^2}{2}\sinh^{-1}\frac{x}{a}$

18. $\int \sqrt{a^2-x^2}\,dx = \frac{x}{2}\sqrt{a^2-x^2} + \frac{a^2}{2}\sin^{-1}\frac{x}{a}$